EXHIBIT 11

PUBLIC VERSION -
CONFIDENTIAL MATERIAL OMITTED IN FULL

EXHIBIT 12

PUBLIC VERSION -
CONFIDENTIAL MATERIAL OMITTED

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF DELAWARE

RYANAIR DAC,	)	
	)	
Plaintiff,	)	
	)	
v.	)	Case No.
	)	1:20-CV-01191-WCB
BOOKING HOLDINGS INC.,	)	
BOOKING.COM B.V., KAYAK	)	
SOFTWARE CORPORATION,	)	
PRICELINE.COM LLC, and AGODA	)	
COMPANY PTE. LTD.,	)	
	)	
Defendants.	)	
_____	)	

HIGHLY CONFIDENTIAL PURSUANT TO THE PROTECTIVE ORDER

Video-recorded videoconference deposition of TIMOTHY J. O'NEIL-DUNNE, taken remotely on behalf of the Plaintiff, beginning at 9:03 a.m. and ending at 5:24 p.m., on Tuesday, September 26, 2023, before JOANNA B. BROWN, Certified Shorthand Reporter No. 8570, RPR, CRR, RMR.

APPEARANCES

FOR THE PLAINTIFF:
HOLLAND & KNIGHT LLP
BY: CYNTHIA A. GIERHART, ESQ.
800 17th Street NW, Suite 1100
Washington, DC 20011
(202) 569-5416
cindy.gierhart@hklaw.com

FOR THE DEFENDANTS:

COOLEY, LLP
BY: ALEXANDER J. KASNER, ESQ.
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004
(202) 842-7800
akasner@cooley.com
COOLEY, LLP
BY: KATHLEEN R. HARTNETT, ESQ.
KRISTINE FORDERER, ESQ.
3 Embarcadero Center, 20th Floor
San Francisco, California 94111-4004
(415) 305-6527
khartnett@cooley.com
kforderer@cooley.com

ALSO PRESENT:

THERESA MAJERS, VIDEOGRAPHER

INDEX

EXHIBITS

EXHIBITS

UNANSWERED QUESTIONS
PAGE LINE
(None.)

Remotely; Tuesday, September 26, 2023
9:03 a.m.

(TIMOTHY J. O'NEIL-DUNNE, deponent, was sworn and examined and testified as follows:)

THE VIDEOGRAPHER: We are now on the record. This begins Media File No. 1 in the deposition of Timothy O'Neil-Dunne. We are taking this deposition remotely via Zoom in the matter of Ryanair DAC versus Booking Holdings, Inc., et al., filed in the U.S. District Court for the District of Delaware, Case No. 1:20-CV-01191-WCB.

Today is September 26, 2023. The time is 9:03 a.m. I am Theresa Majers, the videographer, from Magna Legal Services. The court reporter is Joanna Brown also with Magna Legal Services.

Counsel, please introduce yourselves and who you represent starting with the questioning attorney.

MS. GIERHART: Cynthia Gierhart on behalf of Ryanair.

MR. KASNER: Alex Kasner of Cooley, LLP, on behalf of defendants. I have with me Kathleen Hartnett also of Cooley, LLP, on behalf of defendants,

have a technical ability to distribute their products through GDSs for a typical impediment. There were other airlines who chose not to participate even though they could do so.

Q And did those airlines have content available through some computer system where a travel agent could use some technology to go access it?

MR. KASNER: Object to the form.

THE WITNESS: So pre-Internet, if you had an airline who didn't have access to the GDS and there was no Internet, there was only one -- there were only two other possible forms of access: one was to walk into a physical location, and the other one was by the phone.

BY MS. GIERHART:

Q All right. So I think, then, it's -- it's sort of a post-Internet scenario where a travel agent would use some technology to access an airline's content outside of a GDS or through an agreement with the airline direct?

MR. KASNER: Object to the form.

THE WITNESS: Okay. So there's been -- there's a lot of theory about did the airlines' websites make the Internet very strong, or did the Internet make the airlines' websites very strong?

I think they are symbiotic. So the emergence

of the Internet as a -- as a channel of distribution and the emergence of airlines being able to distribute their content outside of the GDS sort of occurred about the same time.

BY MS. GIERHART:

Q Okay. All right. I think, just to wrap up this Section 4, if we could actually just look at this, on page 5 of your report, actually going up, there's a little summary for each section. So I just want to look at the summary for Section 4. Okay. So that's page 5, the paragraph that starts "First."

So you say "The development of modern airline ticket distribution -- split between direct and indirect distribution can be traced through decades of history."

So I think you are saying here airlines have been selling flights through travel agents and -- for decades. Is that, sort of, the first conclusion?

A Yes. I can wax lyrically about how that occurred; but, yes, that's for airline tickets. Yes, there were decades of history where travel agents sold tickets on behalf of the airlines.

Q And that doesn't distinguish between -- well, you just said selling tickets on behalf of the airline.

So this doesn't involve selling tickets

without the airlines' permission; right?

MR. KASNER: Object to the form.

THE WITNESS: That is correct.

BY MS. GIERHART:

Q Okay. So there's not a history for decades of travel agents selling airlines' tickets without their permission?

MR. KASNER: Object to the form.

THE WITNESS: That's not actually true. There was a very robust gray market where travel agents who did have the right to issue tickets distributed theirs to subagents and had a wider network than their own agents. There's a term used which is called "consolidators" or "consolidation," and consolidators would actually sell to large numbers of travel agents. And the airlines actually liked it because it was a way for them to sell unused or wasting inventory to a broader market that wasn't necessarily fully just a regular travel agency.

BY MS. GIERHART:

Q So it was a bit of off-market selling, but the airlines kind of supported it regardless?

MR. KASNER: Object to the form.

THE WITNESS: Some. I said some. And I want to be very cautious that that -- that marketplace

existed. It existed far more in places like Europe and Asia than it did in the U.S., but it was definitely something that happened.

BY MS. GIERHART:

Q And your report does not make a distinction in this early period of a long history of travel agents selling tickets on behalf of airlines.

Your report doesn't make a distinction between authorized and unauthorized sales; right?

MR. KASNER: Object to the form.

THE WITNESS: What I was addressing in my report -- I think I was very explicit about this -- was really about the U.S. market and about the more conventional, mainstream ways that airline tickets were distributed. So I didn't want to go into that level of detail.

There are many, many different exceptions that are possible in these cases. I have tried to avoid doing that so that we would have a report that was only 25 pages as opposed to one that could be several hundred pages long.

BY MS. GIERHART:

Q Right. Okay. But -- so the -- you are saying you were focusing on the mainstream. Sort of a standard protocol of the time was -- in the

pre-Internet era was for travel agents to work through -- to sell airline tickets through approved methods, which were through GDSs, which had agreements with airlines.

Is that -- that's the standard procedure at that time?

MR. KASNER: Object to the form.

THE WITNESS: That was the way that it actually happened. That would be a good characterization, yes.

BY MS. GIERHART:

Q Okay. Okay. So your report, sort of, focuses on the mainstream, authorized sales of airlines through travel agents?

MR. KASNER: Object to the form.

THE WITNESS: With the caveat that the way that I described this was I focused on the United States market, which, as I said, is different from different countries in different jurisdictions.

BY MS. GIERHART:

Q Right. Okay. But that was the case in the U.S. market?

A The majority, yes. I mean, there were consolidations in the U.S. too, and they are fairly well known. Some of them still exist.

Q I'm sorry. A consolidator?

A Sorry. These -- these companies who sell tickets through different forms rather than the conventional form, there are a number of these forms that now exist for different reasons. There has always been alternative forms, and -- but when we are talking about, as I put in the report, the development of the modern airline ticket system, it was conventionally seen as being airline talked to GDS, talked to travel agent, and, then, customers came in and bought tickets from the travel agent. That was the indirect form of distribution.

Q Okay. Okay. And, then, this, sort of, second part of this paragraph is also, then, talking about what we were just talking about, I think. With direct distribution and indirect distribution, they have existed for at least a half a century with the latter using screen emulation and screen-scraping technology for decades.

So, the second point, we are talking about, sort of, the advent of screen-scraping technology. This was an early form where it was manipulation of content that the travel agents already had access to through their relationship with the GDSs; right?

MR. KASNER: Object to the form.

THE WITNESS: You, kind of, concatenated lots of things. Can you unpack that maybe a little bit?

BY MS. GIERHART:

Q Sure. I think in the -- in the -- in your opinion about screen-scraping technology having been used for decades by travel agents, the early form pre-Internet was limited to manipulating content that travel agents already had from the GDS; is that right?

MR. KASNER: Object to the form.

THE WITNESS: I can answer the question this way: because the travel agents used only the GDS as their primary form of interactive system, therefore, the manipulation of screen emulation was based upon the data that was entirely controlled from the GDS down to the terminal. There was nothing between the two.

When you had a PC and you could emulate the screen, then the possibility started to exist because there was now an intelligence that sat at the remote end of the communication, and that was screen emulation initially, and, then, it went to screen manipulation.

BY MS. GIERHART:

Q Right. Okay. I guess I was just making the distinction, it didn't involve yet going out and getting content from somewhere else?

MR. KASNER: Object to the form.

THE WITNESS: I'll disagree with that because the ability to bring other forms of information, again, pre-Internet was there were physical sources of information that existed. So someone might, for example --

Let me try and use an example for this. I might say that, here in this travel agency, I don't want you to use United Airlines, and that was a directive from the management of the travel agency even though the travel agency was perfectly able to do that.

It was possible through manipulating the screen to eliminate United Airlines out of the displays. That's a possibility. That's a good scenario to illustrate how manipulation of the data that was coming to and from could actually change what was happening at the agents' workstation.

BY MS. GIERHART:

Q Right. I understand that as a removal of content. I just meant obtaining raw data. Where the content comes from was from the GDS. That's all I meant. These early, you know, stages of screen-scraping technology didn't yet involve -- and I think we said because of the example with Russia, or you only sell by phone by GDS. By going in person, there was -- there just wasn't yet a way for travel

agents to get content before the Internet if it didn't --

A Actually, there was.

MR. KASNER: Object to the form.

THE WITNESS: Sorry.

Actually, there was. There's a service on which the GDSs are based, which is called the "OAG." The OAG stands for the "Official Airlines Guide." And the OAG publishes schedules for most airlines whether they participate or not. Therefore, I could have access -- and I, in fact, remember doing this -- where you could bring in the information on dial-up to the OAG source, and you could see that there was, in fact, other airlines who were operating at that particular moment on a particular route, who might not be participating inside a GDS.

And, therefore, you would now be able to say "Ah, I see that that can happen. How do I get access to that" -- "How can I make a booking?"

Well, it will say -- you won't know that from the OAG because the OAG only published schedule data. It didn't publish any inventory data.

And, therefore, I can go and say "Ah. But I can see how this flight exists. Let me see how I can get access to it, probably by the phone."

MS. GIERHART: Okay.

THE WITNESS: In fact, there are ways to manipulate the GDS to actually show what is known as schedule data only -- sorry -- schedule data only, which is possible if you know how to manipulate the GDS.

BY MS. GIERHART:

Q So it would be the equivalent, say, of getting flight itinerary information or route information but not -- there was no ability, you know, to book flights or issue tickets. Then, you said, you'd have to call or, you know --

A Yes.

Q -- use some other method?

MR. KASNER: Object to the form.

THE WITNESS: That's a good way to describe it.

MS. GIERHART: Okay.

THE WITNESS: But I just would caution you about the, no, not possible because there was always somebody somewhere in some garage who was tinkering and trying to do things. So I would just say it was definitely not mainstream. Mainstream was the GDS activity.

BY MS. GIERHART:

Q Right. Right. Okay. All right. I'll start going just a little bit into Section 5, and, then, we can take a break. Okay.

So Section 5 is now -- so, now, we are moving into the "MODERN METHODS FOR BOOKING CONSUMER TRAVEL." It starts on page 9 of your report.

And, yeah, actually, just to clarify because, when I was reading this, I just wanted to make sure, this first paragraph under -- under Section 5 at the very top of the page, it says -- this section begins by discussing, sort of, direct and indirect distribution, which I see, in Section 5, it says -- it then discusses how an airline's business model will drive an airline's decision to encourage and pursue a particular mode of distribution. I think that's Section 6.

Can you confirm if that's what you meant?

A Let me have a quick look. Because it is a comparison to -- when I get into Section 6, I go more into the details, which is the airline distribution preference. And, then, I break that out into the full-service and low-cost carriers. So I try to focus Section 5 on probably the context of indirect distribution and then into the detail because airlines were more important for the context of this report is why I created the whole extra section in 6.

Q Okay. So in this -- in this -- all right. So 5.1, we are talking about direct distribution, and you note in here -- I'm trying to find the exact wording. All right. The second sentence, "This channel gives airlines the most control over their pricing and marketing, and it allows them to collect customer data that can be used to improve their loyalty programs and targeted marketing campaigns."

Why -- we talked about this a little before, but why might an airline want to control their pricing and marketing?

MR. KASNER: Object to the form.

THE WITNESS: For many different reasons. Targeting is, as I've put in here, targeting of products and services and marketing.

The loyalty issue is a big one because, prior to the Internet, loyalty was not as precise a science as it is today. So the target -- the control is something that the airlines wish to exert because, if they feel that they can keep a captive customer, they feel that they can leverage that customer to become more valuable to them as a supplier.

BY MS. GIERHART:

Q Why might an airline want to collect customer data?

talking about OTAs being beneficial for the purpose of comparing flight information, not necessarily booking? But -- so when you are talking about information-sharing, aggregation, is that all just related to comparing flights?

MR. KASNER: Object to the form.

THE WITNESS: In general, yes, but it also applies to other forms of travel information like hotels.

BY MS. GIERHART:

Q Okay. But it's not referring to a benefit to the customer of booking flights through an OTA; right?

A This section is just dealing with the information-sharing and aggregation, those two things brought together.

Q Okay. So -- okay. So you -- I just want to clarify and make sure. So the opinion does not indicate any reason why booking a flight on an OTA is preferable to booking a flight through an airline?

A If you look at the penultimate paragraph under 7.2 on page 15, "From a consumer (i.e., traveler) perspective, OTAs can provide a powerful information-gathering and presentation tool, helping consumers by compiling and comparing different flight itinerary and fare information," also known as offers.

So that isn't talking about booking. This particular section does not refer to booking. That is correct.

Q Okay. Okay. And, then, going to the next section regarding screen scraping as an information aggregation tool, so now we get to the question of how do you define "screen scraping" as used in this report?

A Right. So I tried to come up with a succinct element here. So screen scraping -- I actually prefer the term "screen pausing" because, if you think about a screen, going back to our emulation story, when we were dealing with emulation, we were only dealing with characters/text that was on the screen.

When you move to a PC in today's highly dense pixel 4K TV screen and 4K computer screen, 104- -- 1080P or whatever this is, you could be looking at the entire screen and say "Well, am I looking at just the screen with every pixel, every color, or whatever; or am I looking at the information?"

So screen scraping could encompass the whole of that screen including the bits down at the bottom. I just see it on my screen. I've got a weather app. I'm not going to use that because that's not really relevant to us.

Screen pausing is taking data that is relevant

out -- that's why I prefer to use that term -- which is the process of -- it's automating the process of accessing a web service and recording or scraping/pausing the information displayed on the screen.

So, when we are looking at the website, I'm looking at the data as opposed to all of the different pixels.

Q Okay. It seems limited to visual data that you'd see on the website, but, then, you scrape or pull; is that correct?

A Yeah. It's --

MR. KASNER: Object to the form.

THE WITNESS: Sorry. It's the data as opposed to whether it's green screen, green type, white on green, or yellow on black. Yes. It's related to the data that's in there rather than just the pixels.

BY MS. GIERHART:

Q Okay. And in relation to OTAs and airlines referring to screen scraping here, present day, would refer to pulling light itineraries and routes, maybe ancillary options, pulling all of that data?

That's all screen scraping; right?

MR. KASNER: Object to the form.

THE WITNESS: So screen scraping this, and as

I put in here, the process extracts data from a computer screen by reading the pixels, starting at that level, on the screen and converting those into the -- and getting rid of the unnecessary stuff into text or other data formats that can then be used.

So, in this particular case, you described a series of elements that one can find on an airline website such as flight information, times, or whatever it might be and pulling that into and then making that available.

BY MS. GIERHART:

Q Okay. And would you call it something different, the process of -- of OTAs actually booking the flights?

Connecting to the API and booking a flight, is that called -- that's not screen scraping, but that's something else?

That's booking a flight; is that accurate?

MR. KASNER: Object to the form.

THE WITNESS: So consider that you -- we are doing two different things here. So the first thing, screen scraping, per se, is just grabbing the data, bringing it into an area that makes it usable. That's screen scraping.

When you are doing something else like you are

manipulating the data or doing something with it as I describe in some of the elements below, then that's different. Then I'm now processing the information in the same way that I would if I was a human on the machine.

BY MS. GIERHART:

Q Well, there's processing information, and, then, there's also -- or there's also connecting to an API to book a flight.

Would you consider that part of the screen-scraping process or something else separate?

MR. KASNER: Objection to the form. Testifying. Foundation.

THE WITNESS: So I think what I've described is you need to get a hold of information before you can do something with it. Once I've got the information into a work -- let's call it a work area -- then I can do something with that data.

So, if I'm going to use it and redisplay it, that's an action. Then the screen-scraping effort is done. I haven't manipulated the data, or I might have manipulated the data and put it somewhere, but I'm not interacting with the place where I got the data.

MS. GIERHART: Okay. I think I understand.

THE WITNESS: See, an API -- when you describe

an API, an API is essentially -- the key term is the "I." That's the interface. So an application programming interface means it's bidirectional. So I can get stuff, and I can put stuff.

When I'm screen scraping, screen pausing, I'm grabbing the data. I'm only doing one of those functions, whereas, an API, you have to do both.

BY MS. GIERHART:

Q Right. Okay. And so I think -- I just wanted that understanding going into, then, the next part of the report. For example, it talks about screen scraping offers several key benefits. So I just wanted to understand that when you are saying screen scraping offers benefits, we are talking about going and pulling data from airlines and displaying -- well, maybe not even displaying it. That might be a Step 2.

But the act of pulling data from an airline's website is what you are talking about when you -- when you say "Screen scraping offers several key benefits"; is that right?

A That's correct.

Q Okay.

A In fact, if I can, sort of, look at the title of 7.2.1, which is screen scraping as an information aggregation tool --

Q Uh-huh.

A -- grabbing the information from different sources using screen scraping as one of the techniques, that's what I'm trying to get across.

Q Okay. You go through the key benefits.

Did you consider the downsides of scraping in forming your opinion?

A Yes.

MR. KASNER: Objection to the form. Foundation.

BY MS. GIERHART:

Q Okay. What downsides did you consider?

A There's many different downsides -- two potential downsides. Whether or not those downsides are particularly important or not depends on the use. So I think it sort of depends. The question, I think, overall is to consider the benefits in total as aggregate -- sorry -- not aggregate but unbalanced.

Is screen scraping a useful capability to bring information in that can then be manipulated by the human or by a machine?

Screen scraping is a very common tool for doing that, and I described that.

But, yes, I did consider alternatives, which is not just there are downsides to it, which can be

varied.

Q Okay. I think while -- I'm going to go to Exhibit -- the next exhibit, which is your article referenced in Footnote 14, the second one, the Han and Anderson article --

A Uh-huh.

Q -- which is going to be Exhibit 129.

(Deposition Exhibit 129 was electronically received and marked for identification.)

BY MS. GIERHART:

Q Okay. Do you see that?

A I'm there. Where would you like me to go?

Q Okay. Can you go to page 102 of the article.

A Well, this was -- hold on a second because the way it's coming up is I've only got to page 16.

Q Oh, sorry. It's page 14 of 16.

A Okay.

Q It says 102 at the top.

A Oh, I see. Yes. Understood.

Q Okay. So this page of the article talks about "Ethical Concerns."

Do you see that?

A I do.

Q Actually, sorry to make you jump. I should just go to the title of the article so we know -- have

some context. So the article is entitled "Web Scraping for Hospitality Research: Overview, Opportunities, and Implications." So, in that context, we are talking about web scraping and ethical concerns related to it.

Do you remember reading this when you were writing your report?

A Yes. And -- I do.

Q Okay. Do you see, at the bottom of the first paragraph there in that section, it says: "However, web scraping information that is accessible exclusively to the members and requires logging in is illegal, as this behavior explicitly violates the terms of service"?

A Hold on a second. I'm sorry. I can't find that. I remember the issue. I just can't see it. Where is it? We are on the same page?

Q It's in the middle of the page on the left-hand --

A Oh, yeah, I see it. So -- yes. "However, web scraping information that is accessible exclusively," yes, I see it there.

Q Do you agree with this?

A So I went to look at the original Supreme Court ruling. My interpretation of it -- see, I'm not a lawyer -- said that, in general, it's --

there is legalized screen scraping; right?

So the question of this particular question -- this particular issue about whether this behavior violates the terms of service or not, I think, is an interpretation that these guys who wrote the article had. And since web scraping -- screen scraping is such a common behavior, if this was the case, I think there would be a lot more cases where this behavior was deemed to be illegal because this is not the only place that does it. Web scraping occurs in many other industries, and I think I cite other articles where it takes place. Travel just happens to be one of them because --

Q I -- I was just going to say I think the distinction here is just mentioning scraping where there's a login and violating the terms of service.

So I'm just wondering if that's a determination that you considered or not in writing your report.

MR. KASNER: Objection to the form and testifying.

THE WITNESS: Yes, I did.

BY MS. GIERHART:

Q Okay. What did you consider about that?

A I considered several things in here, one of

which was whether or not it's required. So I looked at actually going in to Ryanair and seeing what Ryanair did on its website. And, in fact, I show -- at the back of my report, there are some sample screens that show that I went in. And, with my assistant, we actually looked at two scenarios where we went in using Ryanair direct and where we went in using Booking.com to see if there was really a difference between the two.

Q A difference in what?

A In how we got to the data and whether the customer was impacted at the end of that.

Q Okay. And when you refer to getting the data, so we are talking about anything leading up to booking the ticket but not including booking the ticket?

MR. KASNER: Object to the form.

THE WITNESS: That is correct.

BY MS. GIERHART:

Q Okay. Actually, did your report take an opinion at all regarding anything at the moment of booking a ticket or after booking the ticket?

MR. KASNER: Object to the form.

THE WITNESS: I think we described what happens in the screenshots to show that we actually went through that, and we talk about myRyanair as part

of the -- it's in the report.

BY MS. GIERHART:

Q Okay. I guess I didn't phrase it very well. I guess, just generally, your report, when it talks about benefits of screen scraping or aggregating data, OTAs being able to show different flights, that's really all talking about the search process leading up to but not including booking a ticket; is that right?

MR. KASNER: Object to the form.

THE WITNESS: Most of the piece is concerned with web scraping, which is grabbing information. Remember, that's a one-way kind of way to look at it. That's pulling the information in. That's not interacting with it out.

BY MS. GIERHART:

Q Okay. Okay. All right. So I just want to look at another -- on the same page, but if you move to the right column, it's about five lines down. The author is saying: "A common ethical concern regarding web scraping is related to the problem of sending too many requests to the host over a short span of time" and that [as read] "A typical web scraper involves querying a website repeatedly. If overused, the practice can prevent others from accessing the website."

Are you familiar with this practice?
A I am very familiar with this practice.
Q Okay. And what -- how are you familiar with it?
A The behavior of lots of requests coming in is a common challenge that a website has to deal with. So, for whatever reason, floods of requests come into a website, and they can be driven by a wide range of different reasons. It could be I'm running a promotion, which could stimulate traffic. It could be that there is something known as a DDOS, a denial-of-service attack. It could be that somebody is trying to scrape my site if I'm the site itself.
So having been on the side of someone who has been the target of scraping, you build tools to support and ensure that you don't have it, and you hope that whoever is on the other end, who is intending to scrape you, is behaving ethically -- sorry. Ethically is a judgment decision -- is behaving well so that you don't actually do things like that to prevent a customer from coming in. I'm familiar with this practice.
Q Is it a known, I guess, harm to airlines' websites --
Is web scraping known to harm airlines' websites?

MR. KASNER: Object to the form.
THE WITNESS: It's known to be capable to harm airlines' websites. I chuckle at this because, in the early days of the Web, it was a real problem. Now we have tools which identify this type of request, and if you are a well-managed website, you can figure out ways to eliminate that so that you don't have to deal with it whether these unnecessary requests are coming in. So you can monitor it. You can determine it. Those tools exist and are well deployed by every competent website owner.
BY MS. GIERHART:
Q Okay. Would you agree there's some expense in developing those tools and constantly fighting scraping activities?
MR. KASNER: Object to the form.
THE WITNESS: I have more stories I can tell you about having to do it, and most, if not all, website owners, whether it's airline or otherwise, should expect this behavior. And most of the website tools now give you a way to handle it, and it's cost-effective.
BY MS. GIERHART:
Q And did you consider these downsides when forming your opinion in this report?

A Yes. And my experience definitely points me in that direction. I've been -- I've been at this -- a victim, if you like, of a website I was responsible for, receiving lots and lots of requests from both scrapers and from bad actors.
Q Which air -- which website was that?
A I believe I alluded to this earlier when we were running the earlier lines, and we had a portion of the website on each one of seven airlines. So our particular section was the one getting hammered.
Q Sorry. Was this -- I missed it.
Was this Flair Airlines or with a different --
A No. And one of the things in my CV, it talks about a company called Air -- excuse me -- Air Black Box, and Air Black Box was the technology provider to a series of airlines. There were seven of them in all. And we ran a portion of the website that created a particular function, and we ran that function for each one of the airlines. So we didn't just have to deal with one airline's website. We actually dealt with seven and an eighth, which was our own, which was ValueAirlines.com. So we were very familiar with these types of behavior.
Q Is there a reason why you didn't include these disadvantages in your report and weigh them and then

ultimately conclude that, you know, one outweighs the other? because they are just absent from the report.
MR. KASNER: Object to the form. Testifying.
THE WITNESS: I felt that the best way to talk about this was to have other people describe it. I've got -- as I said, I've got personal experience of doing this. From this experience and my understanding of how airlines work, it's standard practice that you expect this type of behavior to occur. So I didn't think it was necessary to go beyond that because you have an article such as this which describes those types of conditions.
BY MS. GIERHART:
Q Right. I mean, you just have to do a lot of digging to go find it, and there's no indication in your report that you are going to --
You didn't say "See disadvantages in Footnote 14." It just says "screen scraping is a process of accessing a web service. See Footnote 14."
MR. KASNER: Objection to the form. Argumentative.
THE WITNESS: I think the purpose of quoting this article as I've tried to do, in other words, is there is a lot of information available to explain to a user -- I'm sorry -- a reader that these issues do

exist. I hoped that by quoting esteemed journals like this particular one of the value that that brought. Cornell Hospitality School is very well known, and this publication is pretty authoritative; right?

I don't have to agree with everything, but it's a good layout of exactly what happens. I mean, they are talking mostly about hospitality sites, but it does apply to airline sites. And I felt this was a good way to illustrate the challenge by quoting something from an authoritative journal.

MR. KASNER: Cindy, at the end of this line of questioning, can we take a break?

MS. GIERHART: I think we can do a quick break now and then come back. Another ten minutes?

MR. KASNER: Okay.

THE WITNESS: Okay.

THE VIDEOGRAPHER: We are off the record at 3:52 p.m.

(Off the record.)

THE VIDEOGRAPHER: This is the beginning of Media File No. 6. We are back on the record at 4:02 p.m.

BY MS. GIERHART:

Q Okay. So we were just talking about -- in the report, you mention the benefits of screen scraping,

and you indicated that you also considered the disadvantages of screen scraping. I want to come back to that for a second, if we can switch gears, and I wanted to look at some of your blog posts of yours.

So I can share -- this is going to be Exhibit 130. You might have to zoom in a little bit.

(Deposition Exhibit 130 was electronically received and marked for identification.)

BY MS. GIERHART:

Q First, just out of curiosity, I notice your blog is called -- and your Twitter handle is Professor Sabena. Where did that come from?

A So I tend to have a somewhat professorial way that I speak to people in situations. So my nickname was "Professor." And when I was trying to come up with a handle, I wanted to have the antithesis of something that might appeal to people who were aviation, in other words, AV geeks and people who knew about the airlines, and Sabena was probably the world's worst airline. In fact, its acronym stands for "such a bloody experience, never again." So Professor Sabena is my nickname, my nom de plume, as it were. This -- I see this is from what I wrote in 2008.

Q Yes, it is from a while ago, but it is talking about "Ryanair clamps down on Screen Scrapers." So it

still seems relevant surprisingly to this day.

MR. KASNER: Objection to the form. Testifying.

BY MS. GIERHART:

Q Do you remember this article -- or this blog post?

A I remember writing it. I don't remember a lot of the detail. So there's a lot in here.

Q Okay. Well, first, it starts off saying Ryanair is going after screen scrapers. So the second paragraph there says "So it is not surprising that occasionally they go after intermediaries who are re-selling the Ryanair product."

So I guess, first, the question is just, it was known in the industry since at least 2008 that Ryanair didn't want screen scrapers on their website; is that right?

MR. KASNER: Object to the form.

And if you need a second to read the article, please do.

THE WITNESS: Yeah. Let me read the whole article.

MS. GIERHART: Yeah. Just let me know when you are done.

THE WITNESS: Yeah. Yeah, I remember writing

this. I'm finished.

BY MS. GIERHART:

Q Okay. So, to start, it seemed, in the industry, it was known that Ryanair did not approve of screen scrapers; is that right?

MR. KASNER: Objection to the form.

THE WITNESS: Sorry. It's been well known for a long time that Ryanair does not like screen scrapers, and it is well known that other airlines don't like screen scraping.

BY MS. GIERHART:

Q Okay. And why would they not like screen scraping?

A So in this article, which, remember, is very old -- it is 15 years old. That's a lifetime, maybe generations inside the travel industry, inside the Web for sure. So, at the time, this was something that was, I think, more of a challenge. It was still a fairly new phenomenon at the time about screen scraping. So screen scraping was not -- was fairly novel at this time. I think that's the best way you can describe it.

Q Okay. And as far as why -- why would airlines at this time not like screen scraping?

A I actually put in here that there was actually

some risk. Where is the bit in here? I read that. It says -- if you go to the paragraph, it says "I have an issue with the use of Screen Scraping" --

Q Uh-huh.

A Right? -- "and have never been a huge fan despite the value that can be derived"; right?

[As read] "The core root of the problem is that the backend systems are really not designed to cope with these types of searches and therefore it creates a resource issue which can ultimately cost money. It is fair therefore for a scraper to" -- sorry. "Is it fair therefore for a scraper to make money out of somebody else's resources?"

So I asked that question, a rhetorical question, and I don't answer it. I just pose it as a question.

What is interesting is, at the time, any form of screen scraping -- remember, this is long before big DDS attacks had become prevalent. So it was just a resource issue because there wasn't the tools of sophistication.

Now, I think that it is completely different, and this is normal behavior to expect, screen scrapers or something particularly in travel sites to come and hit you. So the tools have been developed. So I don't

think that statement is true anymore.

Q So it's your belief that screen scraping no longer causes harm to an airline's website?

MR. KASNER: Object to the form.

THE WITNESS: I'm not saying that. I'm saying that it is not as much of an issue as it was when I wrote this article.

BY MS. GIERHART:

Q Okay. And what is that based on?

Have you lately worked with airlines to combat screen scraping?

MR. KASNER: Object to the form.

THE WITNESS: I would characterize it in a slightly different way, is that when I last had to deal with it in my own personal experience and in my discussions with airlines who now have to deal with the problem on their websites and other travel websites because it's not just airlines who get screen-scraped, it is also -- sorry -- OTAs who get it. In fact, I talk there about one website called "Trip.com."

At the time, Trip.com was owned by somebody else, and they were a bad actor. But, now, it's owned by Ctrip. It's become a good actor in that they're using Trip.com in a good way.

So I think it's part of the standard behavior

inside the industry that you expect that people will use some form of screen-scraping tool, and therefore, I think it's kind of a nonissue these days. Whereas before, when I wrote this article 15 years ago, it absolutely was a big issue.

BY MS. GIERHART:

Q You were starting to say -- you were comparing from when you had to deal with it with conversations you are having with people today.

When did you have to deal with it?

A Oh, sorry. This was in the 2017, '18, and '19.

Q And it was a big problem then?

A No. It was -- it was conquered, if you like, from that time. So seven years after -- 2017. Seven years after this article, by then, I had started Air Black Box. This was from 2010. I started Air Black Box in 2012, and we had to deal with it. And it was one of the first questions I asked of my web team.

And I said, "Okay, how do we defend against this? How do we defend against massive amounts of data requests? How do we synthesize those data requests? How do we stop bad actors from entering?"

And those, by then, even in 2012, have become

a lot more normal because this issue was raised at this time and the people had started to deal with it. By the time we were in public in 2016 with Air Black Box and its customers, it was an absolute normal situation, and we were dealing with this as we were dealing with other things like DDOS attacks.

Q And do you think you were getting screen-scraped at the same level that airlines like Ryanair are?

MR. KASNER: Object to the form.

THE WITNESS: Can you give me perhaps a better definition of "level."

BY MS. GIERHART:

Q Do you think that the volume of screen scraping that your company received would be equivalent to the volume of screen scraping that Ryanair receives?

MR. KASNER: Object to the form.

THE WITNESS: Ryanair was a massively much larger airline. It's one of the top airlines on the Web anyway. So we were nothing like that. We were confined to a very specific area in Asia Pacific. So, the total volume, I would say no; but in comparison to what I understand Ryanair is getting, I don't have any data on that subject. I tried to find some, but I can't find any. I would say, probably, it was far more

impactful on us than it was on Ryanair.
BY MS. GIERHART:
Q Even though it would be a much smaller volume?
A Yes. Because let's say we got a thousand bookings a day and we had 10,000 hits that came from screen scrapers, that would translate into -- Ryanair into literally millions of hits in comparison.
Q And so --
A We got a lot, and in one particular case, I can recall most of the traffic coming in was from people attempting to scrape, and they were actually a malicious actor in the specific case that I'm talking about because they were trying to create a copy of our website so that they could create a fake website.
So the detailed knowledge was not just that the actor was doing something bad if we count screen scraping as bad; but they had a malicious intention of doing that not to the benefit of our customers but diverting customers to their website so they could steal from consumers. That's what --
Q Okay.
A -- (inaudible) OTAs do.
Q But -- so you don't know the level of attacks that Ryanair has to deal with in the last five years or even today; right?

MR. KASNER: Object to the form. Foundation.
THE WITNESS: I do not have specific knowledge of that, no.
BY MS. GIERHART:
Q Okay. So you can't really know if the problem is solved for Ryanair, can you?
MR. KASNER: Object to the form.
THE WITNESS: I believe there is a document -- I don't recall which one -- which talks about what Ryanair does to combat the issue. I think they even quote it in one of the responses. They quoted the amount of money that they spend on anti-DDOS and screen scraping. It didn't appear to be a lot of money, which is what I would expect for an organization the size of Ryanair.
BY MS. GIERHART:
Q I would be curious which document you are referring to.
MR. KASNER: Object to the form.
THE WITNESS: God, I'd have to go back and look, but it was one of the responses. Do I have to -- can I find it now, or should I just wait?
MS. GIERHART: I'd rather we move on if you don't know offhand.
THE WITNESS: No. I mean, I recall the

document. I recall that it actually talks about what Ryanair provisions as tools, and it has a cost factor associated with it. And, to me, in my professional opinion, the size of the business that Ryanair does and the number of bookings and web searches that they receive, that seemed to be quite normal.
BY MS. GIERHART:
Q Is it a widely held belief in the industry that screen scraping just isn't a problem anymore?
MR. KASNER: Object to the form.
THE WITNESS: Thinking about that, obviously, I can't speak for everybody nor the industry as a whole, but I would say that screen scraping is normal behavior inside the travel industry, and people accept it and handle it.
BY MS. GIERHART:
Q Right. But none of that says that it isn't a problem. Do you agree that there is a difference?
MR. KASNER: Object to the form. Argumentative.
THE WITNESS: I think, if we were to ask a slightly different question, "Do people accept that DDOS attacks are normal?" we would say the industry expects it. So we deal with it. I think the same applies to screen scraping. Although they are very

different reasons, I think we just accept it.
BY MS. GIERHART:
Q I guess the question is you had the opinion in 2008 that you had an issue with the use of screen scraping and you were not a fan.
Are you a fan today of screen scraping?
MR. KASNER: Object to the form. Foundation.
THE WITNESS: I will never say that I am a fan of screen scraping. That's not what I'm saying here. What I'm saying here -- and I think I still hold the belief -- is if there is a better way to do things, then we should go for it. Screen scraping should not be something that should be a be -- a goal in and of its own self; but if it's the only way to achieve the goal, then you just accept it, and you move on. You do what you can.
BY MS. GIERHART:
Q I wanted to go up just a little bit in the article. You talk about there are three main types of screen scraping. Type 1, you say, is aggregate and then forward the user to the supplier. So you give the example of Kayak. So I think that's your example you were talking about where you aggregate data and then link to the supplier.
And, then, Type 2, you say, is "screen scrape

aggregate and then make the booking in the background without redirecting the consumer." Would --

Is Booking.com in the Type 2 category?

A So these types were defined by me in 2008. Since I haven't seen this article in at least 15 years, I'd have to think about that more before I could give a considered opinion, and since you are presenting this to me, I would ask for time to think about that question.

Q Sure. You can think about it.

And your response could be, you know, "In 2008, this is how I would have considered it in 2008."

A In 2008, I think it was clearer because there was really -- there weren't as many options or as many nuances to it. So judging by the three criteria, if somebody else had written this, if I was using that as the criteria, I would say kind of, yes, that they are a Type 2, not completely yes as Type 2 because I would probably have had -- if I was looking at this now, the different types of screen scraping, I would have many more differentiations.

Q Okay. And I guess is the same --

Would Priceline fall into that Type 2?

MR. KASNER: Object to the form.

THE WITNESS: I would answer it the same way.

MS. GIERHART: Okay.

THE WITNESS: And if you look there, I say that -- I name examples here of three -- of two players, V-Tours and Bravofly. And at the time -- and my memory is going to be a bit sketchy here because I haven't thought about this for a while, but Bravofly, if I recall correctly, was a well-known bad actor of Type 2, which is why I called them out because they did not provide any indication to the customer as to what they were doing and they had lots of problems and it was, kind of, well known.

The issue, I think, is the whole industry has moved on; and therefore, it doesn't appear to be the same as the situation that visited in 2008. I think it would be unfair to label either Booking or Priceline or even any other player you care to name in the same category without considering them in much more detail.

So I would just say I tend to think that there were very specific players in 2008, and I named two of them who I think were bad actors at the time. Whether or not they cleaned up their act, I can't say.

Does -- do I think that Booking and Priceline -- Booking.com and Priceline exactly fall into this category? I'd have to think about that.

///

BY MS. GIERHART:

Q Okay. I guess, going into your thinking there, Type 2, two paragraphs down, you say "Type 2 is the nasty one because it is essentially fooling the customer."

You do mention Ryanair here. Then you also say "It is also EXPRESSLY verboten with Ryanair and just about everyone else." And, then, you go into the terms and conditions for Ryanair.

A Correct.

Q So it sounds -- was your belief in 2008 that this Type 2 where the screen scraper is making the booking on the website without redirecting, that those were, sort of, this -- I mean, you clearly had strong feelings against them; is that accurate?

MR. KASNER: Object to the form. Misstates the document.

THE WITNESS: I'd like to agree with counsel. I think you are misstating the point here. What I am trying to say here is that -- that for the Type 2 -- I tried to focus very clearly on Type 2, and I was expressly thinking about these two players, V-tours, who I don't think exists anymore, and Bravofly that was, kind of, a cross-seller [ph].

And Bravofly was well known for basically

hoodwinking the customer, not just in the area of screen scraping but in other areas, and they acquired a very bad reputation for doing that inside the industry.

So my first categorization is, yes, if you are completely fooling the customer, which I believe that Ryanair -- I'm sorry. Excuse me. Not Ryanair. Excuse me -- that Bravofly was. And I don't think the same thing applies, which is why I think that Booking.com and Priceline don't fit the mold for the strict interpretation I had at the time with Type 2.

Again, this is reinterpreting something that I had from 15 years ago.

BY MS. GIERHART:

Q Okay. I guess, to scroll down to the paragraph we were looking at before where it starts off "I have an issue with" and it ends with that you said there's a rhetorical question, is it fair, therefore, for a scraper to make money out of someone else's resources?

Do you have an opinion on that?

MR. KASNER: Object to the form.

THE WITNESS: I think that's still valid. That's still a valid statement because, again, going to the specific case that we are talking about, Bravofly, which is the one I remember, Bravofly was absolutely

making money and was doing -- I think Michael O'Leary calls them "pirates," the pirate behavior of charging a completely different price to the one that Ryanair was putting out.

So the issue was different, and I think, therefore, that the statement is true that, in this particular case, the scraper was making money by misrepresenting things to the consumer and using the resources of the airline because they were charging a significant premium.

BY MS. GIERHART:

Q Okay. So if an OTA substantially marked up an airline's flights and was able to sell that flight through screen scraping through using the airline's resources, that would not be fair in your opinion?

MR. KASNER: Object -- object to the form. Calls for speculation.

Can I answer?

MR. KASNER: Yes.

THE WITNESS: So I would add one component, and was not being clear, as was the case with Bravofly, purporting to be something that they were not.

BY MS. GIERHART:

Q Okay. I guess, just in that case, what were they not being -- because I'm not familiar with it,

what were they not being clear to the customer about?

A I -- again, this is 15-year-old memory, and a lot of water has flowed under a lot of bridges. I just remember Bravofly was well known for misrepresenting to consumers what it did, and it wasn't just an issue of screen scraping for which they actually did a lot of screen scraping, not just to Ryanair, but to a bunch of other players. And it was well known at the time that Bravofly -- again, this is recalling from memory -- that Bravofly was a bad actor. And I think that when you look -- put all of those things together, I think my statement was accurate --

Q Okay.

A -- for Bravofly.

Q And, then, at the end -- okay. Well, actually, I just want to go to the very end of this, but to go back up to the start to orient what this article -- the catalyst for the article, in the second paragraph, you say this is about the stories that Ryanair has stated, effectively Monday, they will automatically cancel any bookings made by screen scrapers. I think you mean screen scrapers.

A I did.

Q Okay. So with that context, just to go to the very bottom, you conclude, the second to the last

paragraph, "In my humble opinion therefore Ryanair is well within its rights to take this action even though it is objectionable and puts the consumer in the hot seat." Do you still hold that opinion today?

MR. KASNER: Object to the form.

THE WITNESS: I think that if you take the case of a bad actor, that's if Ryanair was the good actor here and the person being damaged and the bad actor was someone like a Bravofly, I'd say yes.

BY MS. GIERHART:

Q And why is that?

A For the reason I think we've already discussed and for the reasons we have just gone through.

MS. GIERHART: All right. I'm going to show you one more blog post. Let me make sure -- okay. It's going to be Exhibit 131.

(Deposition Exhibit 131 was electronically received and marked for identification.)

BY MS. GIERHART:

Q It's from a similar time frame; in fact, I think only a week later. You can have a minute to read through it.

A I can't see it yet.

Q Oh. I'm sorry. Exhibit 131, do you have that?

A I don't have it. Oh, sorry. There it is. Yeah, it came up. Sorry.

Q No worries.

A Well, this is from August, I believe. The other one was from June. Was it from June? Oh, no. This is a week later. You are absolutely right. Okay. Let me just read this one.

Q Sure.

A I don't know if I can because it's a pdf. I was going to say, can we go and look at the article in Aviation Week? I doubt that we would be able to get to it --

Q Yeah. I'm not --

A -- because that would give context.

So I think I'm pretty clear here. Again, it's ten years ago. I was saying the bad Type 2s are actually the ones causing the problem. I'm trying to distinguish between bad Type 2s and good Type 2s are actually the ones causing the problem here.

Ryanair is well within the rights to go after these guys at the time in the jurisdiction that this happened. If, however, Ryanair went after Type 1 scrapers, then I would be less enthusiastic. By implication, the good Type 2s would also fall under that category.

the same as what a conventional travel agency does.
You need to have this minimum information in order to make a reservation. Yes, that's true. You can then use that information and look -- when you combine that with, for example, the behavior of the customer. So you can get -- you do get personalization out of this. That's standard business.
BY MS. GIERHART:
Q Okay. And, then, your next sentence is that, as such, Ryanair is attempting to establish itself as a competitor of online travel agencies.
Is it your contention that Ryanair is trying to establish itself as a competitor because it collects the name, the date of birth, and contact information of its customers?
MR. KASNER: Object to the form. It's one of the things that Ryanair does. Ryanair wants to provide the same level of service as an OTA does because, you know, we've gone through this at great length, looking at comparing OTAs and direct airlines, and I think Ryanair is trying to do the same. They want to be competitive with online travel agencies in gathering the customers' eyeballs.
BY MS. GIERHART:
Q You are saying they have the same business

model, which is to collect information about customers and then target marketing to them; right?
MR. KASNER: Object to the form.
THE WITNESS: That is one of the things they do, yes.
BY MS. GIERHART:
Q Isn't that what essentially every company that markets to consumers today does?
MR. KASNER: Object to the form.
THE WITNESS: That's no different. You are absolutely correct.
BY MS. GIERHART:
Q So is -- wouldn't that make every company a competitor of OTAs?
MR. KASNER: Object to the form.
THE WITNESS: Let's put it this way: Amazon today -- touch wood -- does not sell travel on Amazon.com. So I think the context of this is very important. You can't do a search from Dublin to Stansted on Amazon.com. You can on an OTA, and you can on Ryanair.com. So I think the context is very important.
BY MS. GIERHART:
Q Okay. Is there anything else supporting your contention that Ryanair is a competitor with OTAs other

than what's here in the report?
A Sorry. Ask the question again.
Q Is there anything else that supports your contention that Ryanair is attempting to establish itself as a competitor of OTAs other than what you have here in the report, which is that they are both targeting customers?
MR. KASNER: Object to the form.
THE WITNESS: That is a very broad way that Ryanair and OTAs are competitive, and they both consider each other to be competitive. So, in my report, I've tried to demonstrate how you get to that point with what is the behavior of an airline, what is the behavior of intermediaries be they GDS or OTAs or other forms, and what is the behavior of the customer.
So, as a broad brush, I tried to give that -- the answer to that information so that, inside the report, I tried to give a way that describes how Ryanair behaves, and that's very similar to how an OTA behaves, which is what's good.
MR. KASNER: I think we are at time. So, if we could just wrap this up, that would be great.
MS. GIERHART: Right. No. Yeah. I think -- I think I'm -- we are at time. So I think I'm done.
MR. KASNER: Okay.

THE WITNESS: Okay.
MS. GIERHART: Did you have any questions, or are we done?
MR. KASNER: We do not have any questions. We would like to designate the transcript for this as highly confidential pursuant to the protective order.
THE VIDEOGRAPHER: Joanna, do you want copy orders on the record?
THE REPORTER: Yes, please. Does anyone need a copy of the transcript today?
MS. GIERHART: Yes. We'll -- e-tran only. We don't need paper copies, and if we can, get the rough and expedited.
MR. KASNER: The same for us would be great.
THE REPORTER: Thank you.
THE VIDEOGRAPHER: All right. This concludes the deposition on September 26, 2023, at 5:24 p.m.
(Deposition session concluded at 5:24 p.m.)
-oOo-

I certify (or declare) under penalty of perjury under the laws of the State of California that the foregoing is true and correct.

Executed at _____ on _____.
(Place) (Date)

_____
(Signature of Deponent)

DEPOSITION OFFICER'S CERTIFICATE

STATE OF CALIFORNIA)
) ss.
COUNTY OF ORANGE)

I, Joanna B. Brown, hereby certify:

I am a duly qualified Certified Shorthand Reporter in the State of California, holder of Certificate Number CSR 8570 issued by the Court Reporters Board of California and which is in full force and effect. (Fed. R. Civ. P. 28(a)).

I am authorized to administer oaths or affirmations pursuant to California Code of Civil Procedure, Section 2093(b) and prior to being examined, the witness was first duly sworn by me. (Fed R. Civ. P. 28(a), 30(f)(1)).

I am not a relative or employee or attorney or counsel of any of the parties, nor am I a relative or employee of such attorney or counsel, nor am I financially interested in this action. (Fed R. Civ. P. 28).

I am the deposition officer that stenographically recorded the testimony in the foregoing deposition, and the foregoing transcript is a true record of the testimony given by the witness. (Fed. R. Civ. P. 30(f)(1)).

Before completion of the deposition, review of the transcript [ ] was [ ] was not requested. If requested, any changes made by the deponent (and provided to the reporter) during the period allowed, are appended hereto. (Fed. R. Civ. P. 30(e)).

Dated: _____

_____

DEPOSITION OF: TIMOTHY JAMES O'NEIL-DUNNE
DATE OF DEPOSITION: Tuesday, September 26, 2023
CASE: *Ryanair DAC v. Booking Holdings Inc., et al*, Case No. 20-1191-WCB

ERRATA SHEET

The following are the corrections which I have made to my deposition transcript:

Pg.	Ln.	Now Reads	Should Read	Reason
9	12	"(inaudible)"	gist	Transcription
21	21	pr?cis	précis	Transcription
25	14	buying	enabling	Correction
34	11	crew	credentialed	Transcription
58	18	airline	airline,	Transcription
65	19	serve	settle	Correction
81	2	electronic	paper	Correction
102	11	DDS's	GDSs'	Transcription
109	4	obstructed	abstracted	Transcription
112	12	bot	board	Transcription
113	17	Dog	dot	Transcription
141	10	educators	all comers	Transcription
148	9	part	path	Transcription
149	16	office	offers	Transcription
152	7	gain	against	Transcription
153	21	NDCs	NDC	Transcription
156	21	selling	sourcing	Transcription
158	17	agency or a regular brick…	agency or via a regular brick…	Transcription
160	16	Saudi	Saudia	Transcription

DEPOSITION OF: TIMOTHY JAMES O'NEIL-DUNNE
DATE OF DEPOSITION: Tuesday, September 26, 2023
CASE: *Ryanair DAC v. Booking Holdings Inc., et al*, Case No. 20-1191-WCB

Pg.	Ln.	Now Reads	Should Read	Reason
160	16	Saudi	Saudia	Transcription
161	21	indirect	direct	Transcription
167	19	reaching	reach and	Transcription
168	19	Flair,	Flair then,	Transcription
173	19	that	it	Transcription
179	13-14	"— like the"	"— like the – "	Transcription
190	20	airline's website	airlines' websites	Transcription
217	2	Sendai	Cendyne	Transcription
217	3	Sendai	Cendyne	Transcription
219	10	pausing	parsing	Transcription
219	25	pausing	parsing	Transcription
219	15	104	1040	Transcription
220	4	scraping/pausing	scraping/parsing	Transcription
220	21	light	flight	Transcription
223	5	pausing	parsing	Transcription
230	11	DDOS	DDoS	Transcription
232	8	earlier line	Value Alliance	Transcription
232	22	ValueAirlines.com	www.ValueAlliance.com	Transcription
238	19	DDS	DDoS	Transcription
241	6	DDOS	DDoS	Transcription
243	12	DDOS	DDoS	Transcription
244	23	DDOS	DDoS	Transcription

DEPOSITION OF: TIMOTHY JAMES O'NEIL-DUNNE
DATE OF DEPOSITION: Tuesday, September 26, 2023
CASE: *Ryanair DAC v. Booking Holdings Inc., et al*, Case No. 20-1191-WCB

Pg.	Ln.	Now Reads	Should Read	Reason
260	19	customer's water	customer is like water	Transcription
261	21	FAA	2FA	Transcription
267	21	pr?cis	précis	Transcription
277	24	DDOS	DDoS	Transcription
286	16	form. It's one	form." [Insert new line] "THE WITNESS: It's…	Transcription. The transcript improperly combines Mr. Kasner's objection with The Witness's answer
293	Index	DDOS	DDoS	Transcription
307	index	DDS's	GDSs'	Transcription
307	Index	DDS	DDoS	Transcription
307	Index	DDOS	DDoS	Transcription
331	Index	pausing	parsing	Transcription
334	Index	pr?cis	précis	Index should be corrected to reflect "précis"
340	Index	scraping/pausing	scraping/parsing	Transcription

I, the undersigned, declare under penalty of perjury, that I have read the above-referenced deposition transcript and have made any corrections, additions or deletions reflecting my true and correct testimony.

EXECUTED this 15 day of November 2023, at 10:15 P<.

Timothy James O'Neil-Dunne
Timothy James O'Neil-Dunne

EXHIBIT 13

PUBLIC VERSION -
CONFIDENTIAL MATERIAL OMITTED IN FULL

EXHIBIT 14

PUBLIC VERSION -
CONFIDENTIAL MATERIAL OMITTED

Article

Web Scraping for Hospitality Research: Overview, Opportunities, and Implications

Cornell Hospitality Quarterly
2021, Vol. 62(1) 89–104

DOI: 10.1177/1938965520973587
journals.sagepub.com/home/cqx
SAGE

Saram Han[1,2] and Christopher K. Anderson[1]

Abstract

As consumers increasingly research and purchase hospitality and travel services online, new research opportunities have become available to hospitality academics. There is a growing interest in understanding the online travel marketplace among hospitality researchers. Although many researchers have attempted to better understand the online travel market through the use of analytical models, experiments, or survey collection, these studies often fail to capture the full complexity of the market. Academics often rely upon survey data or experiments owing to their ease of collection or potentially to the difficulty in assembling online data. In this study, we hope to equip hospitality researchers with the tools and methods to augment their traditional data sources with the readily available data that consumers use to make their travel choices. In this article, we provide a guideline (and Python code) for how to best collect/scrape publicly available online hotel data. We focus on the collection of online data across numerous platforms, including online travel agents, review sites, and hotel brand sites. We outline some exciting possibilities regarding how these data sources might be utilized, as well as discuss some of the caveats that have to be considered when analyzing online data.

Keywords

web scraping; online review; data collection; Python

Introduction

The internet presents many interesting opportunities for understanding consumer choice. Any information displayed on any travel website represents a potential data source for hospitality researchers. For instance, a hospitality researcher might collect a list of reviews from TripAdvisor and perform text mining, or retrieve satisfaction ratings to perform some statistical analysis, or determine the relationship between price and page ranking at an online travel agent (OTA). Despite the appeal of these opportunities, however, the fact remains that manually collecting a large amount of online data is inefficient and practically impossible. A typical hospitality research study may require data from multiple hotels across different markets over a prolonged period of time; collecting this data manually can be incredibly time consuming and tedious. Therefore, hospitality researchers need to automate the process of gathering and storing the information presented on travel websites. This is where web scraping comes into play. Web scraping is the process of creating a computer program to download, parse, and organize data from the web in an automated manner (vanden Broucke & Baesens, 2018).

As Table 1 illustrates, there are several alternative approaches that hospitality researchers may consider when trying to automate the collection of online data. One promising option is using an Application Programming Interface (API) provided by the firm hosting the data. However, APIs are often difficult to access, accessible only for a limited duration, or require upfront fees even when researchers are able to access an API. Furthermore, APIs generally will not expose all the required variables.

Recently, scholars have utilized alternative approaches to collect consumer behavior data that mimics online behaviors. Previous studies directly captured data in a simulation lab or survey setting where research participants acted as if they were actually completing online activities (Choi et al., 2017; Min et al., 2015; Shin et al., 2019). However, a common criticism of these studies is that there is a mismatch between real customer behavior and the behavior of the subjects in these experiments, which are often conducted by undergraduate students or participants in Amazon's Mechanical Turk (Schahn & Holzer, 1990). Moreover, although it is possible to conduct large-scale

[1]Cornell University, Ithaca, NY, USA
[2]Seoul National University of Science and Technology, Republic of Korea

Corresponding Author:
Saram Han, Seoul National University of Science and Technology
232 Gongneung-ro, Nowon-gu, Seoul 01811, Republic of Korea
Email: saramhan@seoultech.ac.kr

Table 1.
Comparison of Common Data Collection Methods in Online Market Research.

	Scraped Data	Commercial Web Scraping Service	API	Survey
Cost	Low	Medium	Low/Medium	High
Sample frame	Website users	Website users	Website users	Flexible
Customizability of variables	Medium	Low	Low	High
Ease of frequent collection	Easy	Moderate	Easy	Hard
Data type	Behavioral	Behavioral	Behavioral	Attitudinal
Limitations	Time and programming skills	Data may not be suitable to the researcher's need in terms of variables or content	Limited availability	Time and programming skills

Note. API = Application Programming Interface.

survey research, a number of barriers make that process difficult. Meanwhile, the scale of web scraped data usually is large in nature. Compared with conducting a survey, collecting publicly available online data is inexpensive—even considering the required fee for the commercial service. The monetary cost and effort required to conduct a survey increases linearly with each additional sample, whereas web scraping requires only a one-time expense. Hotel prices change daily, and online reviews are posted even more frequently. In such a dynamic market, a single sampling may not be sufficient. Conducting multiple surveys sequentially is expensive and requires a lot of effort, whereas collecting publicly available online data enables researchers to extract as much data as they need in real time.

Another data collection method that researchers frequently chose is to outsource the scraping task to third-party commercial firms (Wu et al., 2015). Although this is perhaps the simplest option, there are potential pitfalls that may arise due to the lack of discretion in the data collection process. For instance, researchers often realize too late that they forgot to ask the firm to scrape for certain variables. As a result, they are forced to either spend additional time, effort, and money or complete their research without having access to all the desired variables or behaviors.

However, the existing literature points out several drawbacks to web scraped data. One common critique of online data is that study results derived from this data are rarely generalizable to the behavior of the entire target population; although these data are useful to researchers who are interested only in the behavior of online customers, if the researcher's target population includes both online and offline customers, any results derived from online data will be highly vulnerable to selection bias. This is because web scraped data may overrepresent the unique behaviors of online customers. Meanwhile, researchers who opt to utilize surveys as their research method have the flexibility to design a sample frame capable of representing both online and offline customers. Another drawback of publicly available data is that researchers have little discretion in measuring aspects from the sample. In contrast, researchers can include anything they want in a survey or questionnaire, and thus have no restriction in the variables that they can collect. Researchers should carefully decide whether web scraping is suitable to their research and fully consider the potential challenges web scraping poses. An ideal approach might be to collect data through both experiments and web scraping to confirm both the internal and external validity (Viglia & Dolnicar, 2020). For example, some researchers use web scraped data to develop their hypotheses and use traditional data collection methods to confirm these hypotheses (Kupor & Tormala, 2018). Although web scraping has become popular recently and presents an important opportunity to better understand the online marketplace, traditional data collection methods remain the most commonly used strategies in hospitality research.

Table 2 illustrates the potential benefits of web scraped data for hospitality research by summarizing papers by data type with a focus on online reviews in travel marketplaces. To gather this data, we surveyed six highly reputable hospitality journals using the advanced search feature that is available on any journal website. Our analysis included all papers written prior to July 2020. We classified papers as having utilized traditional data collections methods (such as surveys or interviews) if they included the term "online review" and at least one of the terms {"interview," "survey," "amazon mechanical turk," "questionnaire"}, but none of the terms {"scrape," "scraping," "crawl," "actual review," "python"}. In contrast, any papers including the term "online review" and at least one of the following terms {"scrape," "scraping," "crawl," "actual review," "python"} were classified as having directly used web data for online review research. The search queries we used are as follows:

- Interview or survey: "online review" AND ("interview" OR "survey" OR "amazon mechanical turk" OR "questionnaire") NOT ("scrape" OR "scraping" OR "crawl" OR "actual review" OR "python").

Table 2.
Data Collection Methods Used for Online Review Research in Hospitality Journals.

	Interview OR Survey	Web Data
Cornell Hospitality Quarterly	16	3
Journal of Travel Research	21	4
Journal of Hospitality and Tourism Research	7	3
International Journal of Hospitality Management	138	21
Annals of Tourism Research	31	5
Tourism Management	82	25

- Web data: "online review" AND ("scrape" OR "scraping" OR "crawl" OR "actual review" OR "python").

As the table indicates, only a handful of papers use data from the internet to tackle research questions that are related to online marketplaces.

The increasing availability of open source web scraping tools has made it a lot easier for the researchers to build their own customized web scrapers. This article aims to dismantle some of the barriers that hospitality researchers encounter when attempting to utilize web scraping in their online studies. However, this article is not a complete guide to web scraping, but rather an introduction to some of the key tools and requirements for scraping key hospitality data sources. Our article provides hospitality researchers with tools and techniques for collecting data from typical, interactive hotel websites. Due to the introductory nature of this article, we focus on applied concepts and functional illustrations. In this article, we assume that readers have fundamental knowledge of programming languages, such as defining variables, loops, and functions, but not necessarily Python. Proficient programmers and researchers who are familiar with web scraping may already be capable of writing their own code from scratch, and thus may find this article to be too applied or specific. The methods we suggest may not represent the most efficient methods available, as this article only considers websites that have dynamic contents—that is, webpages designed to change in response to human interactions (Massimino, 2016): for example, say a researcher wants to scrape not only all the customer reviews on a particular hotel webpage on TripAdvisor but also the profile information from each individual reviewer. TripAdvisor's dynamic website is designed to present detailed profile information only when a user positions their cursor over the reviewer's profile picture. This is where the dynamic-website specialized code that we cover in this article becomes useful. In contrast, if the website contains only static contents—meaning that the website does not change until the user (i.e., client) moves on to another webpage—our dynamic-website specialized code may take longer to function than static-website specialized code. Therefore, for efficient coding, some researchers may want to refer to Massimino (2016) if the website contains only static contents. While the dynamic-website specialized codes can be used to analyze both static and dynamic contents, static-website specialized codes cannot analyze dynamic contents. As the goal of this article is to provide hospitality researchers with a generally useful tool that can be easily applied to a variety of websites, we only consider the dynamic-website specialized code in this article. Popular Python libraries specialized for static website scraping are Requests, BeautifulSoup, lxml, and Scrapy. Those who are interested in detailed instructions from utilizing these libraries and other web scraping methods that can be applied to broader websites should read vanden Broucke and Baesens (2018). We hope our article will make online data collection easier for hospitality researchers and spark greater interest in web scraping techniques. Therefore, our purposes include the following:

- Providing tools applicable to major hotel platforms (i.e., TripAdvisor, Expedia, Marriott.com, and Airbnb),
- Introducing the process in the simplest terms possible, and
- Discussing both the benefits and limitations of analyzing web scraped data.

This article is a step-by-step guide to web scraping, with a focus on websites that are particularly useful to hospitality researchers. Other papers tend to either focus too narrowly on specific websites or discuss web scraping at an abstract level. Considering that different travel websites provide different insights, hospitality researchers rarely utilize only one specific website in their research. Narrowly focused articles require hospitality researchers to learn different web scraping methods for each individual website included in their research. However, overly abstract articles often fail to provide all-in-one, generalized information that novices can apply to their own research. To narrow the gap between the needs of hospitality researchers and the currently available resources, our article focuses on two aspects that have not previously been thoroughly examined in a single paper.

First, rather than introducing different tools for each individual website, we introduce more general tools that are applicable to all major travel websites. Second, after reading our article, researchers will have all the knowledge they need to start web scraping, as our article provides the entire scripts used for scraping every major travel website.

The remainder of this article is constructed as follows. First, we introduce the environments required for running a Python-based scraper. Then, we explain the sample web scraping code for collecting online hotel reviews and prices. In addition to providing samples of key code required for scraping through the paper, in the supplemental appendix we provide complete scripts for scraping reviews and prices from major OTAs, review sites and hotel brand sites. As most web scraped data are secondary data, with the data generation process outside the researcher's control, biases have to be handled with caution. Therefore, we also illustrate some possible biases that researchers should be aware of before analyzing scraped data. Finally, we discuss the academic implications of web scraping and the ethical issues that hospitality researchers should consider when collecting online data for their own research.

Web Scraping Fundamentals

In the following section, we outline the fundamentals of web scraping illustrated through the use of Python—a general-purpose programming language. Our article is not an introduction to Python as a whole, but rather focuses on the aspects of Python that are key to web scraping. Python is a very approachable and intuitive programming language and is often the language of choice for many introductory programming courses. There are many online resources that provide an overview of Python. For those who are interested in learning the basics of Python language, we recommend reading Downey (2014). However, our focus in this section is on the unique methods and skills necessary for scraping hospitality data from a variety of sources using Python. As there is a large scraping community that uses Python3, throughout the article we utilize Python3 specifically (https://www.python.org/downloads/).

Required Environment and Python Codes

As our goal is to simplify the process of gathering online information and transforming it into a meaningful data set, we focus only on the essential elements of web scraping in this article. Over the course of this article, we use three different tools to make a scraper that facilitates the collection of data from major hotel websites: Selenium, WebDriver, and XPath. Figure 1 presents the flowchart of the steps of how we utilize the tools specialized for web scraping the dynamic contents.

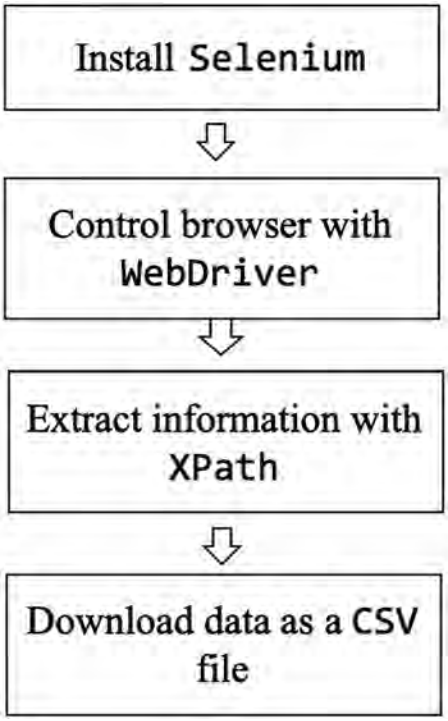

Figure 1.
Flowchart of Web Scraping for Extracting Dynamic Contents.

Selenium. In this section, we introduce the Python library Selenium. Although there are many libraries in Python that facilitate web scraping, Selenium is the most useful when dealing with interactive, JavaScript-heavy pages like those on travel sites such as TripAdvisor, Airbnb, and Expedia. We start by illustrating scraping approaches for extracting review data from TripAdvisor (Listing 1). We then extend this approach to extract review data from other platforms and later extend this approach further to collect other types of data, for example, prices. Let's begin by scraping reviews and basic reviewer profile information from the following TripAdvisor link: https://www.tripadvisor.com/Hotel_Review-g60763-d93344-Reviews-The_Watson_Hotel-New_York_City_New_York.html.

Listing 1.
Define Our Target Website for Scraping.

```
1. target_page = "https://www.tripadvisor.com/Hotel_Review-g60763d93344-Reviews-The_Watson_Hotel-New_York_City_New_York.html"
```

First, let's name this target page. With other Python libraries (BeautifulSoup, request, Scrapy, etc.), you can

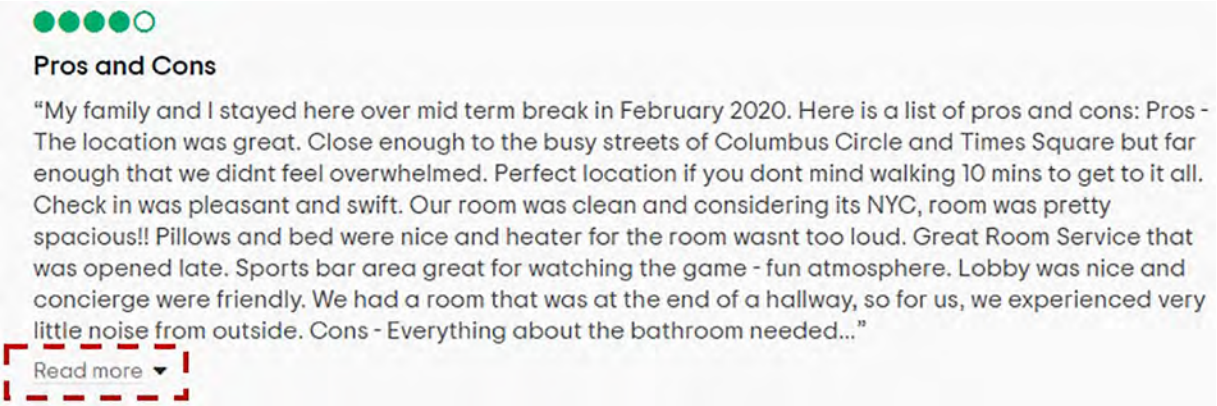

Figure 2.
Example of an Interactive Platform.

scrape everything that is displayed on the screen. While this is good news, it also means that you cannot scrape information that is hidden until you click on it. For instance, contents that are hidden behind the button "Read more," as illustrated in Figure 2, cannot be scraped with other popular Python libraries such as BeautifulSoup, request, or Scrapy. Unlike websites that are only written in Hypertext Markup Language (HTML) or CSS, many interactive travel sites use JavaScript, which allow for interactive functionality on the web page. Contrary to many other programming languages, the core functionality of JavaScript lies in making webpages more interactive and dynamic. Therefore, for sites that make heavy use of JavaScript, we need to write our code in a way that emulates human browsing behavior. That is where the Python library Selenium comes into play. You can easily install it with the following code in the command line: `pip install -U selenium` from the terminal window on your computer. Selenium requires a third-party software called a `WebDriver` which we discuss in the next paragraph.

WebDriver. WebDrivers exist for most modern internet browsers, including Chrome, Firefox, Safari, and Internet Explorer. When using these browsers, a browser window will open up on your screen and perform the actions specified in your code. We can easily download a WebDriver from https://sites.google.com/a/chromium.org/chromedriver/downloads. As WebDrivers exist for most popular browsers, you can choose to download whichever WebDriver best works for you. In the following section, we assume that readers will download the WebDriver for Google's Chrome browser—chromedriver. Make sure you download the file that matches both your operating system (i.e., Windows, Mac, or Linux) and the version of your current browser (i.e., Chrome, Internet Explorer, Edge, or Firefox). The ZIP file you download will contain an executable called chromedriver.exe on Windows, or simply chromedriver otherwise. While locating the WebDriver in the same directory as your Python scripts is the easiest way to call the driver, it is also possible to explicitly pass the location of where you have the WebDriver as we do with the variable `dirver_loc` in the code Listing 2. Once the WebDriver opens the browser, you can interact with the browser until you close it (i.e., `driver.close()`).

Listing 2.
Python Code for Opening and Closing Your WebDriver.

```
1. driver_loc = 'C:/Users/Desktop/chrome
   driver.exe'
2. driver = webdriver.Chrome(driver_loc)
3. driver.get(target_page)
4. #Write your scraper here.
5. driver.close()
```

XPath. Once a page is loaded with WebDriver, you will want to extract the information displayed on the screen and download it on your computer. To extract this information, you first need to view the source code of the website elements you want to scrape. For Windows users, the easiest way to check this information is to right-click on the part within the web page you want to scrape and select "Inspect." You will then see a tree-based view of the HTML codes. Every web page is made up of a bunch of these HTML tags denoting each type of content on the page. By clicking on the arrows, you can see the nested structure of the code. To scrape hotel review websites, we only need to understand the basic structure of HTML language, which consists of tag, attribute, and the value of the attribute. Figure 3 illustrates part of the source code underlying the page displayed after selecting "Inspect."

As an illustration, let us assume that the element of the HTML code written to display the reviewer's name (i.e.,

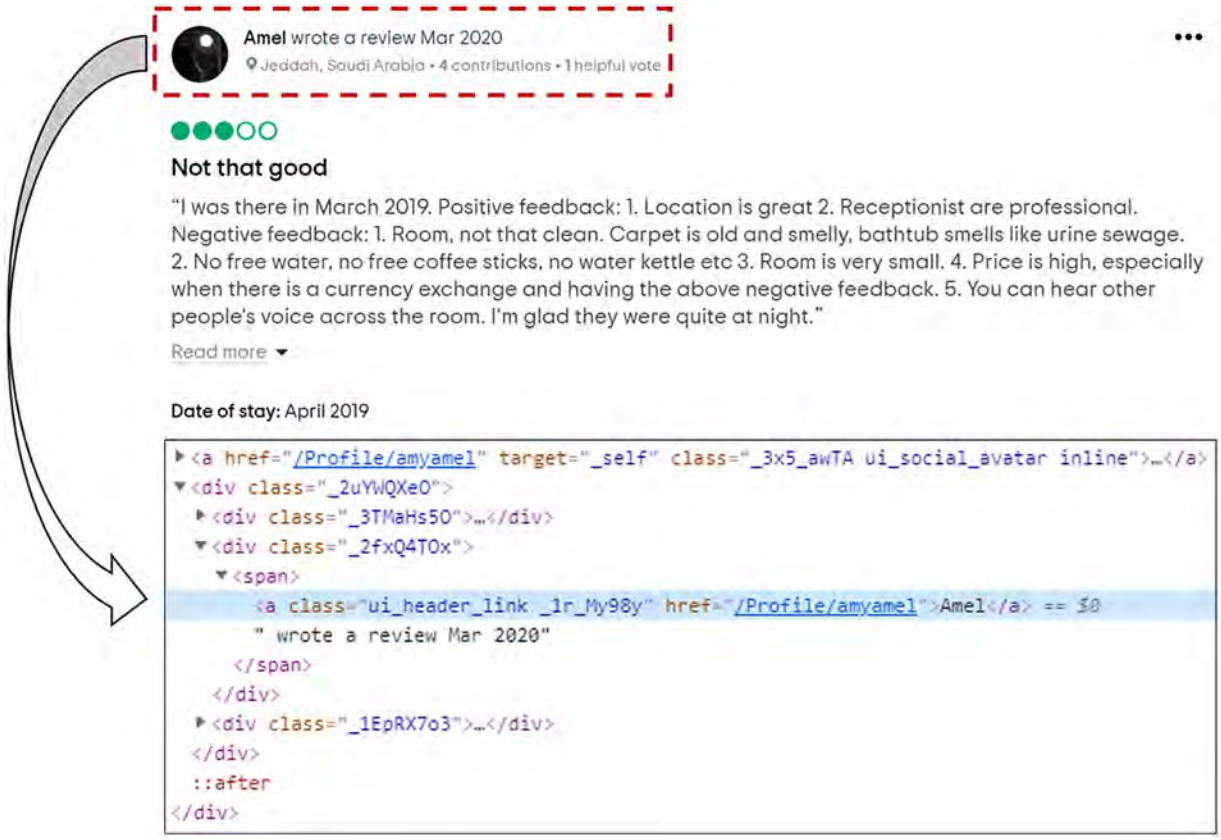

Figure 3.
Displayed Tree-Based HTML Code View after Clicking on Inspect.

Amel) is shown in Listing 3, where the `<a>. . .</a>` is just one of the many existing tags (e.g., `<p>`, `<div>`, `<ul>`) in HTML which encloses the hyperlink and styles the web designer decided to use when displaying the reviewer's name. The value of the attribute "`MemberBlock`" is the name of a style that is assigned to the attribute class. Likewise, `href` is another attribute of this element that redirects users to the reviewer's profile page when they click on the reviewer's username. XPath is a language used to find the location of any element on a web page using this structure of the HTML. Selenium can use the XPath language to select elements. Although there are many Selenium methods, we use only find_element_by_xpath and find_elements_by_xpath throughout this article for ease of application.

Listing 3.
HTML Element for the Reviewer's Name.

```
1. <a class="MemberBlock" href="/Profile/amyamel">Amel</a>
```

To extract the information from the HTML code Listing 3, the XPath has to be written as, `//a[@class='MemberBlock']`, which basically is `//Tag[@Attribute='Value']` that can be interpreted as "grab the current node (`//`) with the tag name where the class attribute corresponds to `MemberBlock`." However, oftentimes the value of the attribute is very long and contains multiple special characters. This can be difficult because if even one of the characters is missing, a serious error may occur. `//a[contains(@class, 'Member')]` presents an easier way to call the same element, by utilizing the same process but not requiring users to include the entire value of the attribute. This XPath prevents the researcher from provoking an error simply by not indicating the exact value of the attribute. The function `contains()` enables Selenium to easily find the desired element, even if the researcher only provides a fragment of the attribute value. Accordingly, the `find_element_by_xpath` method in the code Listing 4 indicates the element of the HTML code Listing 3.

Listing 4.
Indicating the HTML Element.

```
1. driver.find_element_by_xpath(".//a[contains(@class, 'Member')]")
```

To extract the reviewer's name, we can add .text as shown in the code Listing 5.

Listing 5.
Indicating the Text Content.

```
1. driver.find_elements_by_xpath(".//a
   [contains(@class, 'Member')]").text
```

Note that the value of the same attribute can always change (e.g., `MemberBlock`) whenever the web designer defines a different name for the value. Therefore, the same web scraper may not work if the web designer updates the HTML code. If an error occurs after running the codes that we provided in the supplemental appendix, we recommend our readers to review the HTML code of the target website and check whether the value of the attribute has changed.

If the researcher is interested in extracting the user's profile link, the following `.get_attribute("href")` method is useful for finding the value of the `href` attribute (Listing 6).

Listing 6.
Indicating the Value of the "href" Attribute.

```
1. driver.find_element_by_xpath(".//a
   [contains(@class, 'Member')]").text.
   get_attribute("href")
```

Applications for Major Hotel Platforms

In this section, we apply these scripts to scrape four major hotel review websites and discuss how to best deal with platform-specific features when building a scraper. We also discuss how hospitality researchers can take advantage of the unique opportunities presented by each individual platform.

Reviews on TripAdvisor. TripAdvisor is the largest player in the travel review platforms arena (Wang & Chaudhry, 2018). Therefore, this platform has been used in many marketing and hospitality studies (Chevalier et al., 2018; Wang & Chaudhry, 2018). TripAdvisor has several unique features that make it attractive to researchers. For instance, each individual's previous platform activities are displayed on their profile. This feature allows researchers to take reviewer level heterogeneity into account as they analyze TripAdvisor data (Gao et al., 2017). Therefore, we recommend scraping each reviewer's profile URL to collect individual-level data that may be useful in the future.

Another interesting feature that recent studies have started to focus on is the existence of *retailer-prompted reviews* (Askalidis et al., 2017; Han & Anderson, 2020; Mayzlin et al., 2014). Unlike regular self-motivated reviews, retailer-prompted reviews are those that are posted in response to hotels' email invitations to post reviews online. This feature allows researchers to investigate how rating behavior differs depending on the nature of the review posting process.

Although most scraping tasks can be completed using the aforementioned functions and XPath, there is one problem that requires additional attention. As mentioned in Figure 2, many TripAdvisor reviews are long, meaning that users must click on the "Read more" button to see the full review. Researchers must be able to automatically click this button to scrape entire reviews. For this purpose, Selenium offers a selection of "actions" that can be performed by the browser, such as clicking elements. Fortunately, in TripAdvisor, once you expand one review, the rest of the reviews within the same page expand as well. If for some reason all of the desired reviews do not automatically expand with a single click, a for-loop can be added to expand each review individually (see Listing 8). Once we identify the element where the "Read more" button locates, we can use the `execute_script` method to expand the review. We recommend using try-except to avoid the scraper stopping in situations where there is no expandable review on the page.

Listing 7.
Python Script to Expand the Review Element.

```
1. more = driver.find_elements_by_xpath
   (".//div[contains(@data-test-target,
   'expand')]")
2. try:
3.     driver.execute_script("arguments[0].
       click();", more[0])
4. except:
5.     pass
```

Listing 8.
Python Script to Expand Every Review Element Within a for-loop.

```
1. for container in containers:
2.     more = container.find_elements_by_
       xpath(".//div[contains(@class,
       'expand')]")
3.     try:
4.         driver.execute_script("arguments
           [0].click();", more[0])
5.     except:
6.         pass
```

Figure 4.
Pagination.

Figure 5.
Infinite Scrolling.

TripAdvisor presents hotel reviews over multiple separate pages. Once you have successfully scraped the information from the first page of reviews, you may want to move on to the following page to scrape older reviews. This is called pagination (Figure 4; Zhang et al., 2020). We can use the `execute_script` method again to send a JavaScript command to the browser. This method clicks on the "Next page" button, enabling researchers to scrape the same information on the following pages (Listing 9).

Listing 9.
Clicking the Next Page Button.

```
element = driver.find_element_by_xpath('//a[contains(@class, "nav next")]')
driver.execute_script("arguments[0].click();", element)
```

Reviews on Expedia. One big advantage of Expedia is that all reviews are written by valid customers. That is, as Expedia only allows customers who purchased through their website to write reviews, there is a smaller chance that users will see fake reviews written by non-verified customers on Expedia than on other websites such as TripAdvisor, where anyone can write reviews. Moreover, as Expedia sends out email requests to everyone who makes purchases through their website, the entire review collection process is similar to a survey. Therefore, as long as we know who did not respond to the email request sent by Expedia, we can generalize the study results to the population of those who purchased through Expedia. This may not be possible when using data from TripAdvisor, as the TripAdvisor customers' sample frame is unknown.

On Expedia, users may not need to click the "Next page" button as they do on TripAdvisor. Instead, they must scroll down and load older reviews by clicking on the "More reviews" button shown in Figure 5. This is called infinite scrolling. If there is a significant quantity of reviews written about a given hotel, users must scroll further down and repeatedly click on the "More reviews" button until all the reviews have loaded. The scraper must mimic this process. Listing 10 opens the number of reviews that you assign with `revnum` (e.g., we set here as 300). This while-loop continues scrolling down until the accumulated number of loaded reviews exceeds `revnum`.

Listing 10.
Scrolling Down.

```
revnum = 300
loadednum = 0
while loadednum < revnum:
    more = driver.find_element_by_xpath(".//button[contains(@class,'more-reviews-button')]")
    driver.execute_script("arguments[0].click();", more)
    time.sleep(1)
    containers = driver.find_elements_by_xpath(".//div[contains(@class,'uitk-card-separator-bottom')]")
    loadednum = len(containers)
```

Reviews on Airbnb. Airbnb is a peer-to-peer marketplace that emerged as a typical example of what is called the sharing economy. Many studies present evidence that Airbnb threatens the traditional accommodation market system (Zervas et al., 2018). This website is most interesting to researchers hoping to better understand customers' behavior in the sharing economy. Due to the peer-to-peer market nature, previous studies argue that there are unique rating behaviors in Airbnb (Ert & Fleischer, 2019). For instance, in addition to guests rating service providers, guests are also evaluated by the host, and these ratings are made publicly available as well. Owing to this dual review process, suppliers tend to receive overwhelmingly high ratings that are not observed under other hotel review systems (Zervas et al., 2018). Despite these unique features of the sharing economy, little is understood about this system. We hope that, by making Airbnb data more accessible, we can encourage other researchers to further explore the sharing economy.

The scraping code for Airbnb is a combination of the codes used for TripAdvisor and Expedia, as the website randomly changes their review display. That is, when the WebDriver visits the supplier's page on Airbnb, the reviews are randomly listed either as pagination or infinite scrolling. Therefore, once the WebDriver opens Airbnb, we need to first determine which scraper should be applied based on the HTML structure. Although there are many different ways to accomplish this, here we build a `check_exists` function that checks whether Airbnb reviews are displayed in pagination or infinite scrolling.

```
1. def check_exists(self, xpath):
2.     try:
3.         self.find_element_by_xpath
           (xpath)
4.     except NoSuchElementException:
5.         return False
```

After building the function that checks which platform design we have to deal with, we can simply run different scrapers by writing the following if statement.

```
1. # example of xpath that uniquely
   indicates Infinite scrolling
2. xpath = './/a[@class="_16i7snfh"]'
3. if check_exists(driver, xpath)==True:
4.     # Codes for infinite scrolling
5. else:
6.     # Codes for pagination
```

Reviews on hotel brand sites. Major hotel brands have their own websites where customers can write reviews about their experience, just as they can on TripAdvisor or Expedia. Brands offer up their own reviews in an effort to reduce the need for prospective consumers to visit other travel sites such as TripAdvisor or Expedia. Hotel brands are motivated to do this because reservations made on sites such as TripAdvisor and Expedia are more costly to hotels and because customers who visit these sites may end up booking their stay with another company. It is possible that customers who read and write reviews on hotel brand websites differ from those who use other platforms where alternative hotel options are listed. For instance, these customers may be more loyal to the brand or hoping to have their opinions heard by the hotel manager rather than by other potential customers. Platforms like this, where the user groups differ from other platforms users, generate lots of interesting research opportunities.

Although the scraper should be written differently depending on whether the targeted hotel website uses pagination or ultimate scrolling, the overall process of writing the scraper remains exactly the same regardless of the target website. We include an illustration of Marriott.com in Appendix A (along with complete scripts for other travel sites).

Scraping (Prices) at the Market-Level

While scraping online reviews enable researchers to understand customer Word-of-Mouth (WOM) behaviors, scraping prices gives us insight into the market. This requires no major additional functions or methods beyond the scripts that we introduced previously for scraping reviews. The only difference is that we loop over different hotels within a specific market instead of different reviews. Remember that our scraping code differs by whether the page uses pagination or infinite scrolling. While hotel prices listed on Airbnb and TripAdvisor are displayed using pagination, hotels on Expedia are listed using infinite scrolling. Therefore, utilizing the same approach detailed in code Listing 9, we can build our scraper for Airbnb and TripAdvisor to collect price data and click on the "Next" button to move on to the following pages (see code Listing 11).

Listing 11.
Click the Next Button to Collect all Hotel Prices.

```
1. next_exists = check_exists(driver, '//a
   [contains(@aria-label, "Next")]')
2. if next_exists:
3.     element=driver.find_elemet_by_
       xpath('//a[contains(@aria-label,
       "Next")]')
4.     driver.execute_script("arguments[0].
       click();", element)
```

In contrast, for Expedia we design the code to scroll down until there are no additional listings before documenting the hotel prices (see code Listing 11). This approach is equivalent to how we scraped the reviews from Expedia in code Listing 12.

Listing 12.
Scrolling Down to Collect all Hotel Prices.

```
1. while check_exists(driver, ".//button
   [contains(@data-stid, 'show-more
   -results')]"):
2.     more = driver.find_element_by_xpath
       (".//button[contains(@datastid,
       'show-more-results')]")
```

Table 3.
The Number and the Average Length of Reviews Across Three Platforms During 2018–2020.

	TripAdvisor	Expedia	Hotel Brand
Number of reviews	713	2,089	1,760
Average review length	731.12	119.56	141.63

```
3.      driver.execute_script("arguments[0].
        click();", more)
4.      time.sleep(1)
5.      containers = driver.find_elements_
        by_xpath(".//div[contains(@class,
        'link-container')]")
```

Platform Differences Causing Biases

Although collecting and analyzing online review data broadens our understanding, it is important to mention a few relevant caveats. Unfortunately, many studies conducting research using data from online travel sites ignore the biases that arise when using secondary data. Multiple selection biases exist in scraped data, which can impact a researcher's ability to draw insights about the target population (i.e., all customers who stayed at the hotel). In this section, we present some of the noteworthy differences between hotel review platforms that hospitality researchers should be aware of before collecting and using web scraped data.

Review Differences Across Platforms

Owing to differences in how reviews are collected and written, collected reviews may vary in terms of number, valence, and content depending on what hotel review platform they were posted on (Litvin & Sobel, 2019). This is true even for reviews that were written about the same hotel and during the same time period. This is true for a couple of reasons. First, every hotel review platform has its own objectives and its own review collection process. To encourage review submission, Expedia sends customers a post-stay email with a link to submit a hotel review. TripAdvisor, however, relies fully on customers' self-motivation to post reviews: unless individual hotels choose to partner with TripAdvisor to invite their verified customers to post reviews, either by sending an email or through their online reputation management firms (e.g., ReviewPro, Revinate, and Medallia). These differences in the review collection process among these platforms yield systematic differences in terms of the review scores and their content. Second, customers may perceive different platforms as having different audiences. Therefore, different types of customers may prefer different platforms depending on who they hope to reach with their review. While customers may perceive distributors' websites (i.e., a hotel's own website or Expedia) as a suitable channel for contacting hotel managers, TripAdvisor might be perceived as an appropriate method for reaching other potential customers. As a result, the dominant valence and topics may differ across platforms.

Several noteworthy patterns are frequently observed across travel platforms. First, a major difference between review platforms is the number of reviews per hotel. To illustrate how a given hotel's quality might be evaluated differently by each of the three platforms, we scraped the reviews of the New York Marriott Marquis, which has a large number of reviews on TripAdvisor, Expedia, and Marriott's own website. For the purposes of this demonstration, we utilized reviews written between October 2018 and December 2019. Table 3 demonstrates that significantly more reviews were posted on Expedia and the hotel's own website than on TripAdvisor. This pattern is particularly interesting because anyone can post a review on TripAdvisor, whereas only verified guests can post reviews on OTAs and hotel brand sites. A major driver in creating this disparity in the number of reviews across the platforms stems from how the reviews are collected. Expedia and hotels encourage customers to post reviews online by sending post-stay feedback invitation emails. This pattern holds across different time periods, as shown in Figure 6, where we plot review characteristics (number, score, sentiment, and length) on a quarterly basis for 2018 and 2019. The average review length is significantly longer on TripAdvisor than on the other platforms. Self-motivated (i.e., non-email prompted) reviewers are more likely to put greater effort into writing reviews, given that they have already undertaken the effort of opening TripAdvisor and posting a review. Therefore, review length (calculated in number of words) is greatest on TripAdvisor, where most reviews are self-motivated. After eliminating stop words, which refer to the words that are filtered out due to their extremely high or low appearances, we counted the words that appeared on each platform. As is shown in Table 4, the most frequently used words are slightly different depending on the platform. Unlike Expedia and TripAdvisor, where the most frequent word is "room," the hotel's own site reviewers are more focused on the hotel location.

Finally, the distribution of ratings between the two platforms has different patterns, as is shown in Figure 7. The proportion of relatively negative ratings is higher on TripAdvisor than on Expedia or the brand site. Given the cost or effort of posting a review on TripAdvisor, customers may be more likely to post if their expectations have not been met (or have been exceeded), and as a result these

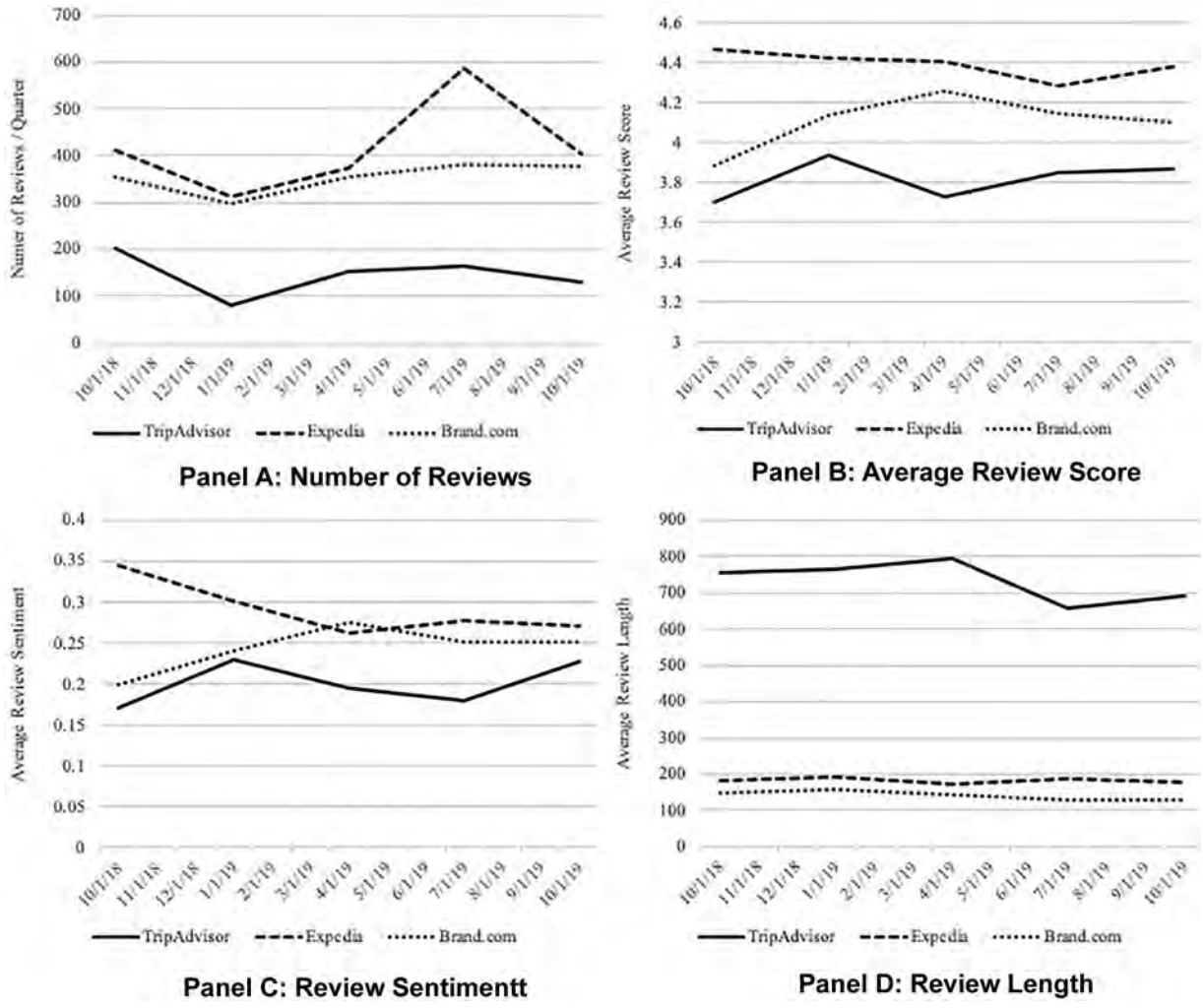

Figure 6.
Differences in Reviews (Number, Score, Sentiment, and Length) by Platform.

Table 4.
Top 10 Most Frequently Used Words for the Same Hotel's Reviews by Platforms.

	TripAdvisor	Expedia	Hotel Brand
1	room (1,346)	room (754)	location (678)
2	time (1,035)	location (691)	room (654)
3	stay (659)	time (482)	stay (644)
4	square (536)	staff (400)	time (562)
5	location (463)	square (334)	staff (488)
6	get (367)	stay (288)	square (352)
7	service (360)	leave (287)	service (295)
8	one (349)	review (282)	excellent (228)
9	staff (336)	comments (281)	good (217)
10	floor (331)	traveler (281)	clean (210)

reviews may be more extreme (and perhaps nonrepresentative; Anderson et al., 1994). In contrast, reviews that were more easily posted (due to the email invitation) are less likely to be affected by the different posting motivation across different ratings. As a result, reviewers that would otherwise have relatively low posting intention (i.e., positive ratings) are able to easily submit their reviews, potentially resulting in higher average ratings at sites with email encouraged/prompted reviews, for example, OTAs and brand sites (Han & Anderson, 2020).

Price Differences Across Platforms

Next to reviews, prices are among the most interesting data types available across different travel platforms. Just as

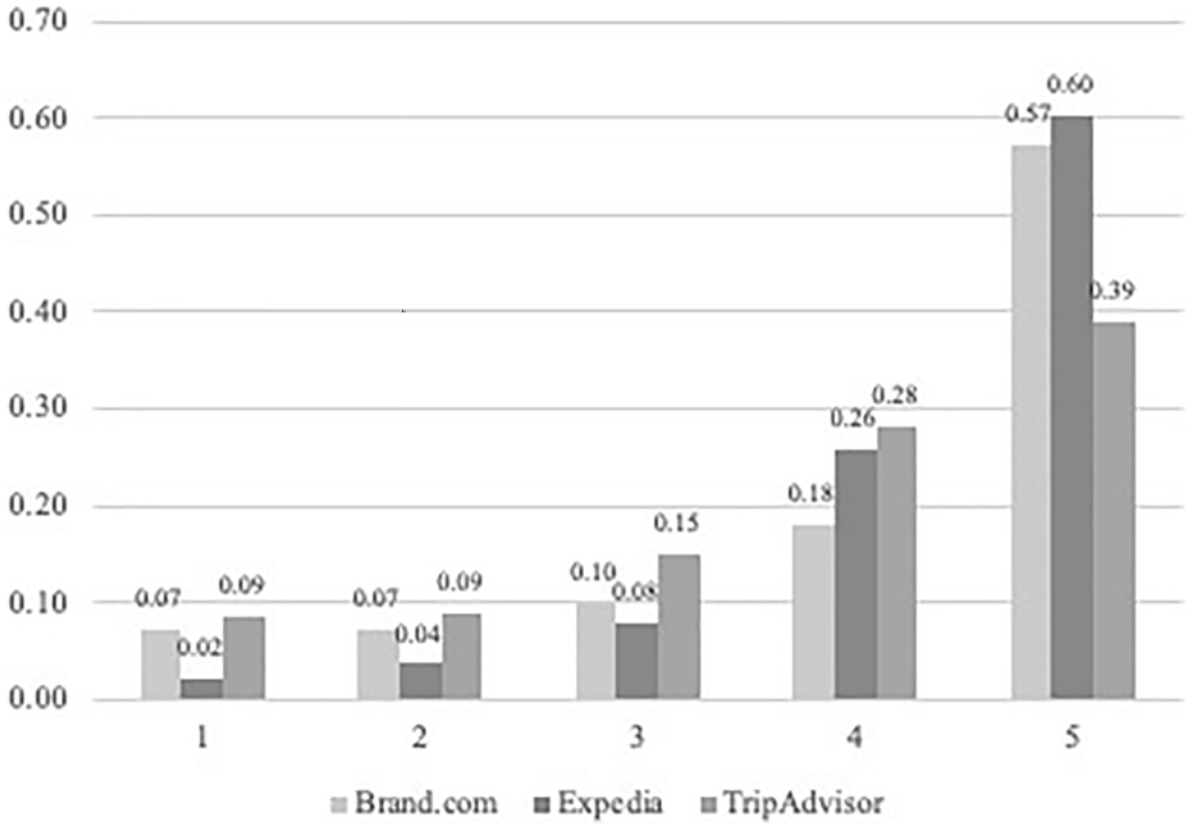

Figure 7.
Distribution of Review Scores by Platform.

Figure 8.
Listing of a Sold-Out Property in Expedia.

with reviews, different platforms have different objectives and different capabilities when presenting prices. As a result, the way in which prices are displayed and communicated across platforms varies, which may influence customers' purchasing decisions. For example, Expedia displays all properties, including those that are sold-out or have limited availability, as is shown in Figure 8. In contrast, other travel platforms, such as Airbnb, only list properties that are available and have price data to be displayed. Customers may behave differently as a result of these differences in information display. Accordingly, it is likely that customers behave differently depending on which travel platform they choose (Park & Jang, 2018). Therefore, assuming that all platforms have the same market structure ignores these differences in information stimuli that influence customer behavior.

OTAs are also considerably more versed in merchandising when displaying price information. Merchandising by OTAs may include such actions as strike-through pricing or scarcity messaging, which attempt to increase conversion

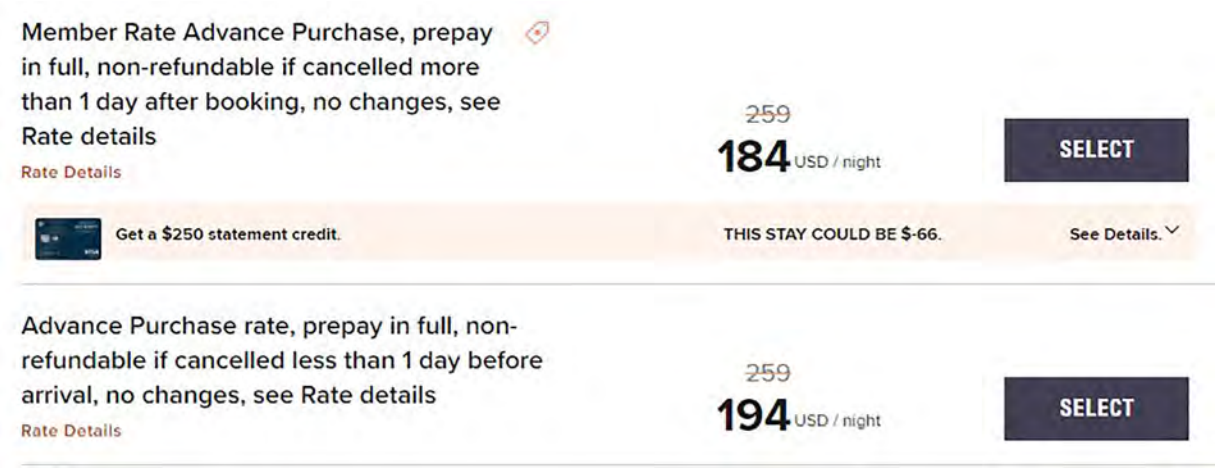

Figure 9.
Different Price Offer to the Members.

rates. Similarly, rate parity has received considerable attention recently. While OTAs want to display prices that can compete with those posted on supplier websites, hotels are often motivated to offer better prices to customers who book directly to avoid OTA commissions. Closed user group (CUG) or membership selling is a common method of offering lower prices to a subset of consumers. In CUG situations, the consumer must login (either at the OTA or hotel site) to access better prices. Figure 9 illustrates strike-through prices and the resulting member rates available to a typical CUG. By definition, CUG discounted prices are only available to those who have logged in; the majority of customers will never see this price disparity. Researchers who analyze the online hotel market using prices scraped from travel websites must understand how these distribution channels influence each other.

Implications

The academic implications of web scraping are twofold: exploratory and confirmatory. In general, web scraped data are suitable for exploratory research where the research question has not been examined in detail. The aim of this research area is to understand the general pattern of the customers in question and find preliminary evidence that warrants a more detailed study. For example, Danescu-Niculescu-Mizil et al. (2013) show that users in online beer communities follow a two-stage life cycle in terms of the language they use: the innovative learning phase and the conservative phase. Wu and Huberman (2010) found that later online product review ratings tend to vary considerably from earlier ones, making overall review ratings less extreme. Although these studies demonstrate interesting behavioral patterns and are valuable in their own right, they still require further testing in an experimental setting.

Web scraped data are also often used in confirmatory studies that test hypotheses from previous studies (Landers et al., 2016; Marres & Weltevrede, 2013). If this hypothesis testing investigates whether one of the variables scraped from the website is an exogenous variable that varies independently of an error term, this study is considered as an experiment (Harrison & List, 2004). Web scraped data enable researchers to conduct experiments in a real environmental context and conduct research without informing research subjects that they are taking part in an experiment. This research design is referred to as a natural experiment. The advantage of a natural experiment is that it accounts for the realistic environment that real customers encounter. Therefore, an ideal natural experiment not only increases external validity but also does so when internal validity is insufficient (Harrison & List, 2004). For example, Han and Anderson (2020) take advantage of a unique characteristic of TripAdvisor, the fact that a portion of its reviewers post after being prompted to do so by hotel managers. By comparing regular self-motivated reviews and prompted reviews, the authors test whether satisfied or dissatisfied customers are more motivated to post online reviews. Web scraped data can also be combined with data from other sources in confirmatory research. Xie et al. (2014) combined TripAdvisor's review data with archival data regarding hotel revenue per available room (RevPAR) matched to the Texas Comptroller's Office database to investigate the impact of various review website attributes on hotel performance.

Managers in the hospitality industry can benefit from web scraping, too. It is well known that customers heavily rely on online reviews before making a purchasing decision

(Brown et al., 2007; Chevalier & Mayzlin, 2006). Therefore, it is important to understand how customers evaluate services. Although there are multiple online reputation management companies that do this job on behalf of firms, hotels can save money by web scraping themselves. For example, firms can web scrape major online review websites in the industry periodically and automate this process, which is referred to as web crawling (Massimino, 2016). Without spending any extra money, firms can obtain a summary of each online review website. Based on these summaries, managers can decide which review websites they should prioritize and invest to maximize the number of reviews they receive.

Ethical Concerns

As legislation on web scraping varies from country to country, researchers should look into local legislation. In this section, we discuss the legal and ethical issues mainly in the U.S. context. On September 9, 2019, the U.S. Supreme Court legalized web scraping in situations where the scraped information is designed to be publicly accessible. The court defined public information as data that are neither available for purchase nor hidden behind a password-protected authentication system. The logic behind the court's decision was that, legally, web scraping is no different than browsing in terms of what data are being requested from a website. However, web scraping information that is accessible exclusively to the members and requires logging in is illegal, as this behavior explicitly violates the terms of service (ToS).

It is also noteworthy that web scraping copyrighted data and re-using them for commercial purposes would be considered illegal. For example, web scraping video contents from YouTube and re-posting them on ones' own website could be illegal as videos are copyrighted.

However, illegally sharing data is not likely a matter of concern for the majority of our target audience: researchers who are interested in using web scraping for academic purposes. In addition, web scraping is a relatively new data collection method in academia, and therefore the law is still evolving (Hillen, 2019). Therefore, researchers must take into account that it is always possible that current laws regarding web scraping will change and that they may need to seek professional legal advice before web scraping.

In addition to legal issues, researchers should also consider ethical constraints when collecting data online for academic purposes. As previous studies argue, legality does not necessarily mean that data usage is entirely ethical (Massimino, 2016). Using online data in research is relatively new in comparison to other data sources. Therefore, its current legality suggests a need for further research regarding the ethicality of web scraped data and the safety of the web scraping practice. By the same token, although web scraping qualifies for Institutional Review Boards (IRBs)' review exemptions in most of the cases (Massimino, 2016), researchers need to be conscientious of any societal entities that may be impacted by web scraping. A common ethical concern regarding web scraping is related to the problem of sending too many requests to the host over a short span of time. A typical web scraper involves querying a website repeatedly. If overused, this practice can prevent others from accessing the website. A web scraper that is written for the purpose of collecting multiple online reviews or hotel prices sends requests to the web server that is hosting the site whenever it opens a new page. While the requests of a human user are usually within a manageable range, a web scraper that makes speedy and bulky automated requests can easily exceed the bandwidth threshold of the host and make the server unresponsive (Massimino, 2016). When the web scraper hits the server with frequent requests, a host may issue a warning or may respond with useless content if web scraping behavior is detected.

Therefore, we recommend inserting a random delay between individual requests, such as limiting requests to three per second (Massimino, 2016). For example, Landers et al. (2016) executed a 2-s delay between each web page request to avoid overburdening the host's server. In addition, scraping during off-peak hours can help reduce the load on the host and increase the speed of the scraping process. Finally, a good practice in web scraping is to carefully read the robots exclusion protocol (REP), which are standardized instructions on whether certain user agents can scrape parts of a website. Typically, this information can be found in the admin page of a website.

Summary

A growing number of customers not only obtain travel information online but also make transactions over the internet. Studies indicate that the internet (as opposed to the offline, voice, or travel agent distribution channels) has become the dominant distribution channel in terms of travel reservations (Park, 2009). Therefore, the importance of studying online customer behaviors cannot be overemphasized. However, there is a strong tendency among hospitality researchers to rely on traditional data collection methods, which are limited in terms of what research questions they can answer as well as the generalizability of their insights. Our work provides a simple method of scraping online reviews and price data using the Python language. As our goal is to make hospitality researchers more comfortable collecting online data, we focus on how the scraping process functions on major travel websites. Although not a comprehensive introduction into Python, we introduce the essential elements necessary to handle and scrape interactive

hospitality platforms such that they can augment readily available introductory Python resources.

Although web scraped data introduce incredible opportunities to hospitality researchers, there are important aspects that must be accounted for when scraping travel websites. As every platform has unique characteristics and purposes, each platform attracts different users, which in turn forms a different market. While ignoring these platform-specific aspects may induce biases in the researchers' analyses, properly making use of these challenges may make platform differences a unique natural experiment setting for exploring new opportunities. At the same time, embracing these differences is what provides for fruitful research. For instance, the social influence effect, which refers to the effect of previous reviews on future reviews, impacts most online review platforms due to the nature of being able to see previous reviewers' opinions. This effect induces potential biases that are not of concern in traditional survey research as there is no chance that survey participants will see other participants' opinions before answering the survey questions. Despite the possibility of inducing social influence bias, Askalidis et al. (2017) overcome this challenge by utilizing reviews written by retailer-prompted reviewers who were invited to contribute their opinions using a separate web page where there is a lesser chance of seeing previous opinions. Going further, they compare this group of reviews to the regular, organic reviews using the difference-in-difference method, and make use of the challenge to identify the social influence effect in the online communities. Another example of turning the challenge into opportunity is the study of Wang and Chaudhry (2018). Many online hotel review platforms allow managers to respond to customer reviews, which may influence future reviewers' opinions. While the managerial response could present a challenge for researchers who want to understand the unbiased opinions of the reviewers, Wang and Chaudhry (2018) identified the managerial response effect by comparing ratings from online review platforms where managerial responses are visible with ratings from platforms where managerial responses are not made visible.

Our article has significant implications for hospitality researchers who hope to better understand the online travel marketplace. We outline simple methods that enable hospitality researchers to collect incredibly useful secondary data that they could not have obtained by relying on traditional data collection methods alone. Although traditional data collection methods are still valuable and, in many cases, cannot be replaced by new methods, online customer behaviors are hardly replicable in offline research design. Even if the purpose of the research is to make a causal inference that can only be tested in a strict lab setting, confirming this effect in the real online marketplace adds value to the study in terms of external validity.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, or publication of this article.

ORCID iDs

Saram Han https://orcid.org/0000-0001-9077-9107
Christopher K. Anderson https://orcid.org/0000-0002-7103-8351

Supplemental Material

Supplemental material for this article is available online.

References

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66.

Askalidis, G., Kim, S. J., & Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems*, *97*, 23–30.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, *21*(3), 2–20. https://doi.org/10.1002/dir.20082

Chevalier, J. A., Dover, Y., & Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*, *37*, 685–853.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Choi, S., Mattila, A. S., Van Hoof, H. B., & Quadri-Felitti, D. (2017). The role of power and incentives in inducing fake reviews in the tourism industry. *Journal of Travel Research*, *56*(8), 975–987.

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No country for old members. *Proceedings of the 22nd International Conference on World Wide Web - WWW '13, Rio de Janeiro, Brazil* (pp. 307–318). Association for Computing Machinery. https://doi.org/10.1145/2488388.2488416

Downey, A. (2014). *Think Python: How to think like a computer scientist*. Green Tea Press.

Ert, E., & Fleischer, A. (2019). The evolution of trust in Airbnb: A case of home rental. *Annals of Tourism Research*, *75*, 279–287.

Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, *95*, 1–11.

Han, S., & Anderson, C. K. (2020). Customer motivation and response bias in online reviews. *Cornell Hospitality*

Quarterly, *61*, 142–153. https://doi.org/10.1177/1938965520902012

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Hillen, J. (2019). Web scraping for food price research. *British Food Journal*, *121*(12), 3350–3361. https://doi.org/10.1108/BFJ-02-2019-0081

Kupor, D., & Tormala, Z. (2018). When moderation fosters persuasion: The persuasive power of deviatory reviews. *Journal of Consumer Research*, *45*, 490–510.

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492.

Litvin, S. W., & Sobel, R. N. (2019). Organic versus solicited hotel TripAdvisor reviews: Measuring their respective characteristics. *Cornell Hospitality Quarterly*, *60*(4), 370–377.

Marres, N., & Weltevrede, E. (2013). Scraping the social?: Issues in live social research. *Journal of Cultural Economy*, *6*(3), 313–335.

Massimino, B. (2016). Accessing online data: Web-crawling and information-scraping techniques to automate the assembly of research data. *Journal of Business Logistics*, *37*(1), 34–42.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, *104*(8), 2421–2455.

Min, H., Lim, Y., & Magnini, V. P. (2015). Factors affecting customer satisfaction in responses to negative online hotel reviews: The impact of empathy, paraphrasing, and speed. *Cornell Hospitality Quarterly*, *56*(2), 223–231.

Park, J. (2009). Consumers' travel website transferring behaviour: Analysis using clickstream data-time, frequency, and spending. *Service Industries Journal*, *29*(10), 1451–1463.

Park, J. Y., & Jang, S. C. S. (2018). The impact of sold-out information on tourist choice decisions. *Journal of Travel and Tourism Marketing*, *35*(5), 622–632.

Schahn, J., & Holzer, E. (1990). Studies of individual environmental concern. *Environment and Behavior*, *22*(6), 767–786.

Shin, H., Perdue, R. R., & Pandelaere, M. (2019). Managing customer reviews for value co-creation: An empowerment theory perspective. *Journal of Travel Research*, *59*, 792–810. https://doi.org/10.1177/0047287519867138

vanden Broucke, S., & Baesens, B. (2018). *Practical web scraping for data science*. Apress.

Viglia, G., & Dolnicar, S. (2020). A review of experiments in tourism and hospitality. *Annals of Tourism Research*, *80*, 102858.

Wang, Y., & Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, *55*(2), 163–177.

Wu, F., & Huberman, B. A. (2010). Opinion formation under costly expression. *ACM Transactions on Intelligent Systems and Technology*, *1*(1), 1–13. https://doi.org/10.1145/1858948.1858953

Wu, L., Mattila, A. S., Wang, C. Y., & Hanks, L. (2015). The impact of power on service customers' willingness to post online reviews. *Journal of Service Research*, *19*(2), 224–238.

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, *43*, 1–12. https://doi.org/10.1016/j.ijhm.2014.07.007

Zervas, G., Proserpio, D., & Byers, J. (2018). A first look at online reputation on Airbnb, where every stay is above average. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.2554500

Zhang, Y., Liu, L., & Ho, S. Y. (2020). How do interruptions affect user contributions on social commerce? *Information Systems Journal*, *30*(3), 535–565.

Author Biographies

Saram Han is an assistant professor at the College of Business and Technology at Seoul National University of Science and Technology. He received his PhD from the Cornell School of Hotel Administration in the Cornell SC Johnson College of Business. His research interests include digital marketing, online reviews, service marketing, and marketing analytics. He earned a BBA in Tourism Management from Kyung-hee University, Seoul, Korea, and an M.S. degree from the Michigan Program in Survey Methodology, University of Michigan.

Christopher K. Anderson is a professor at Cornell SC Johnson College of Business, School of Hotel Administration. He earned his BSc/MSc in engineering from the University of Guelph, and his MBA/PhD from the University of Western Ontario, Richard Ivey School of Business. He teaches and conducts research in data analytics, pricing, distribution, and revenue management.